

IMPLICIT ITERATION METHODS IN HILBERT SCALES UNDER GENERAL SMOOTHNESS CONDITIONS

QINIAN JIN AND ULRICH TAUTENHAHN

ABSTRACT. For solving linear ill-posed problems regularization methods are required when the right hand side is with some noise. In the present paper regularized solutions are obtained by implicit iteration methods in Hilbert scales. By exploiting operator monotonicity of certain functions and interpolation techniques in variable Hilbert scales, we study these methods under general smoothness conditions. Order optimal error bounds are given in case the regularization parameter is chosen either *a priori* or *a posteriori* by the discrepancy principle. For realizing the discrepancy principle, some fast algorithm is proposed which is based on Newton's method applied to some properly transformed equations.

1. Introduction

In this paper we are interested in solving ill-posed problems

$$Ax = y, \quad (1.1)$$

where $A \in \mathcal{L}(X, Y)$ is a linear, injective and bounded operator with non-closed range $\mathcal{R}(A)$ and X, Y are Hilbert spaces with corresponding inner products (\cdot, \cdot) and norms $\|\cdot\|$. Throughout we assume that $y \in \mathcal{R}(A)$ so that (1.1) has a unique solution $x^\dagger \in X$. We further assume that y is unknown and $y^\delta \in Y$ is the available noisy right hand side with

$$\|y - y^\delta\| \leq \delta.$$

In recent literature many aspects of treating ill-posed problems with noisy right hand side have been studied. For an overview, see, e.g., the textbooks [4, 39]. The numerical treatment of ill-posed problems (1.1) with noisy data y^δ requires the application of special regularization methods. In this paper we study *implicit iteration methods in Hilbert scales*, in which regularized solutions x_n^δ are obtained by

$$x_k^\delta = x_{k-1}^\delta - (A^*A + \alpha_k B^{2s})^{-1} A^* (Ax_{k-1}^\delta - y^\delta), \quad k = 1, 2, \dots, n, \quad x_0^\delta = x_0 \quad (1.2)$$

where $B : \mathcal{D}(B) \subset X \rightarrow X$ is some unbounded densely defined self-adjoint strictly positive definite operator, $\alpha_k > 0$ are properly chosen real numbers, s is some generally nonnegative number that controls the smoothness to be introduced into

Date: September 21, 2010.

2000 Mathematics Subject Classification. 47A52, 65F22, 65J20, 65M30.

Key words and phrases. Ill-posed problems, inverse problems, regularization, Hilbert scales, implicit iteration method, order optimal error bounds, general smoothness conditions, operator monotone functions.

the regularization procedure and x_0 is some properly chosen starting value. In these regularization methods, the positive number

$$\sigma_n := \sum_{k=1}^n \frac{1}{\alpha_k} \quad (1.3)$$

plays the role of the regularization parameter. For results on convergence rates of this method in the special case $s = 0$ we recommend the paper [7], and for some extensions to the nonlinear case we recommend [6, 13] and [11, 14, 31].

Method (1.2) with $n = 1$ and $x_0 = 0$ is the method of Tikhonov regularization in Hilbert scales which has been studied by Natterer [28]. From this paper we know that under the assumptions $\|B^{-a}x\| \sim \|Ax\|$ and $\|B^p x^\dagger\| \leq E$ the Tikhonov regularized solution x_α^δ of problem (1.2) with $n = 1$, $x_0 = 0$ and $\alpha_1 = \alpha$ guarantees order optimal error bounds

$$\|x_\alpha^\delta - x^\dagger\| = O(\delta^{p/(a+p)}) \quad \text{for } p \leq 2s + a \quad (1.4)$$

in case α is chosen *a priori* by $\alpha \sim \delta^{2(a+s)/(a+p)}$. This result has been extended

- (i) to the case of choosing α *a posteriori* by the discrepancy principle, see, e. g., [29, 33, 34, 35],
- (ii) to a general regularization scheme, see, e. g., [4, 35],
- (iii) to the case of general source conditions including infinitely smoothing operators A , see, e. g., [10, 17, 21, 22, 25],
- (iv) to the case of nonlinear ill-posed problems, see, e. g., [4, 12, 15, 30, 33, 36].

Our paper is organized as follows. In Section 2 we collect some preliminaries on properties of the implicit iteration methods in Hilbert scales, formulate our general smoothness conditions and give some consequences that follow from the general smoothness conditions by exploiting either operator monotonicity or interpolation in variable Hilbert scales. In Section 3 we treat the case of *a priori parameter choice* of the regularization parameter σ_n and Section 4 treats the case of choosing σ_n *a posteriori* by the discrepancy principle. In Section 5 we discuss practical issues of choosing the starting value x_0 and the parameters s , n and α_k , $k = 1, \dots, n$. In particular, some fast globally convergent algorithm for realizing the discrepancy principle is proposed which is based Newton's method applied to some properly transformed equations. For testing the algorithm, numerical experiments are performed in Section 6.

2. Preliminaries

2.1. Properties of the regularization method. In our further study, instead with B , we will work with the inverse $G = B^{-1}$. Following proposition gives us some equivalent representation for x_n^δ defined by (1.2) along with some preliminary properties which will be useful for deriving order optimal error bounds.

Proposition 2.1. *Let $T = AG^s$, $G = B^{-1}$ and σ_n be defined by (1.3). Then, the regularized solution x_n^δ defined by (1.2) possesses the representation*

$$x_n^\delta - x_0 = G^s g_n(T^*T)T^*(y^\delta - Ax_0) \quad \text{with} \quad g_n(\lambda) = \frac{1}{\lambda} \left(1 - \prod_{k=1}^n \frac{\alpha_k}{\lambda + \alpha_k} \right). \quad (2.1)$$

In addition, the function $g_n : (0, c] \rightarrow (0, \infty)$ with $c = \|T\|^2$ and the corresponding residual function $r_n(\lambda) := 1 - \lambda g_n(\lambda)$ obey the properties

$$(i) \quad g_n(\lambda) \leq \sigma_n, \quad (2.2)$$

$$(ii) \quad \lambda g_n(\lambda) \leq 1, \quad (2.3)$$

$$(iii) \quad \lambda r_n(\lambda) \leq \sigma_n^{-1}, \quad (2.4)$$

$$(iv) \quad r_n(\lambda) \leq \sigma_n^{-1} g_n(\lambda). \quad (2.5)$$

Proof. The proof of the representation (2.1) is standard. For the proof of (i) we follow the paper [7] and observe that the function $r_n(\lambda) = \prod_{k=1}^n \alpha_k / (\lambda + \alpha_k)$ is monotonically decreasing and convex with $r_n(0) = 1$. From these properties we conclude that

$$r_n(\lambda) \geq r_n(0) + r'_n(0)\lambda.$$

Since $r_n(0) = 1$ and $r'_n(0) = -\sigma_n$ we obtain $r_n(\lambda) \geq 1 - \lambda\sigma_n$, which is equivalent to (i). The proof of (ii) follows from the representation $\lambda g_n(\lambda) = 1 - \prod_{k=1}^n \alpha_k / (\lambda + \alpha_k)$. For the proof of (iv), we multiply (2.5) by $\lambda\sigma_n/r_n(\lambda)$ and obtain the equivalent inequality

$$\lambda \sum_{k=1}^n \frac{1}{\alpha_k} \leq \prod_{k=1}^n \left(1 + \frac{\lambda}{\alpha_k}\right) - 1.$$

This inequality, however, always holds true since the left hand side is the first order term of the polynomial in λ on the right hand side. For the proof of (iii) we use (iv) and (ii) and obtain $r_n(\lambda) \leq \sigma_n^{-1} g_n(\lambda) \leq \sigma_n^{-1}/\lambda$, which is equivalent to (iii). \square

Remark 2.2. Note that in our forthcoming analysis we will also exploit the fact that $r_n(\lambda) \leq 1$ for $\lambda \in (0, c]$ which is a consequence of the nonnegativity of $g_n(\lambda)$. Further note that property (iii) of the above proposition tells us that the regularization method (2.1) has at least a qualification of $p_0 = 1$. For the concept of qualification, see [39]. Finally we note that our analysis does not require the full strength of the properties (i) – (iii) of Proposition 2.1. Indeed, property (i) will be exploited for the λ – range $\lambda \leq \sigma_n^{-1}$, and properties (ii) and (iii) will be exploited for the λ – range $\lambda \geq \sigma_n^{-1}$.

For deriving order optimal error bounds for $\|x_n^\delta - x^\dagger\|$ with x_n^δ defined by (2.1) we introduce the regularized solution x_n with exact data by

$$x_n - x_0 = G^s g_n(T^*T)T^*(y - Ax_0).$$

It can easily be checked that the following error representations

$$x_n^\delta - x_n = G^s g_n(T^*T)T^*(y^\delta - y) \quad \text{and} \quad x^\dagger - x_n = G^s r_n(T^*T)G^{-s}(x^\dagger - x_0) \quad (2.6)$$

are valid. From these error representations we see that in the $\|Ax\|$ - norm and in the X_p - norm $\|x\|_p := \|G^{-p}x\|$ we have

$$\|Ax_n^\delta - Ax_n\| = \|Tg_n(T^*T)T^*(y^\delta - y)\|, \quad (2.7)$$

$$\|Ax^\dagger - Ax_n\| = \|Tr_n(T^*T)G^{-s}(x^\dagger - x_0)\|, \quad (2.8)$$

$$\|x_n^\delta - x_n\|_p = \|G^{s-p}g_n(T^*T)T^*(y^\delta - y)\|, \quad (2.9)$$

$$\|x^\dagger - x_n\|_p = \|G^{s-p}r_n(T^*T)G^{-s}(x^\dagger - x_0)\|. \quad (2.10)$$

2.2. Smoothness assumptions. We formulate our smoothness assumptions in terms of some densely defined unbounded selfadjoint strictly positive operator $B : X \rightarrow X$. We introduce a *Hilbert scale* $(X_r)_{r \in \mathbb{R}}$ induced by the operator B which is the completion of $\cap_{k \in \mathbb{R}} \mathcal{D}(B^k)$ with respect to the Hilbert space norm $\|x\|_r = \|B^r x\|$, $r \in \mathbb{R}$. For technical reasons, instead of B we will work with the inverse $G := B^{-1}$, which is a bounded linear injective and selfadjoint operator with non-closed range $\mathcal{R}(G)$. Note that the above Hilbert space norm $\|\cdot\|_r$ may be represented by

$$\|x\|_r = \|B^r x\| = \|G^{-r}x\|, \quad r \in \mathbb{R}.$$

In addition, according to [10, 19] we call a function $\varrho : (0, a] \rightarrow (0, b]$ an *index function* if it is continuous and strictly increasing with $\varrho(0+) = 0$ and assume

Assumption A1. There exist constants $M \geq m > 0$ and some index function $\varrho : (0, a] \rightarrow (0, b]$ with $a = \|G\|$ and $b = \varrho(a)$ such that

- (i) $m \|\varrho(G)x\| \leq \|Ax\|$ for all $x \in X$,
- (ii) $\|Ax\| \leq M \|\varrho(G)x\|$ for all $x \in X$.

Assumption A2. For some positive constants E and p we assume the solution smoothness $x^\dagger - x_0 = G^p v$ with $v \in X$ and $\|v\| \leq E$. That is,

$$x^\dagger \in M_{p,E} = \{x \in X \mid \|x - x_0\|_p \leq E\}.$$

Assumption A1 characterizes the smoothing properties of the operator A relative to the operator G and allows the study of problems with finitely and infinitely smoothing operators A in a unique manner. Typical index functions in applications are power type index functions $\varrho(t) = t^a$ for problems (1.1) with finitely smoothing operators A and index functions $\varrho(t) = \exp(t^{-a})$ where the inverse ϱ^{-1} is of logarithmic type for problems with infinitely smoothing operators A . Such problems appear, e.g., in inverse heat conduction. Assumption A2 characterizes the smoothness of the element $x^\dagger - x_0$ in the Hilbert scale $(X_r)_{r \in \mathbb{R}}$. By using Assumption A2 we can study different smoothness situations for $x^\dagger - x_0$.

Let us give some comment on order optimal convergence rates for identifying x^\dagger from noisy data $y^\delta \in Y$ under the link assumption A1 and the smoothness assumption A2. Let $\mathcal{R} : Y \rightarrow X$ be an arbitrary method and $\mathcal{R}y^\delta$ be an approximate solution for x^\dagger . Then, the quantity

$$\Delta(\delta, \mathcal{R}) = \sup\{\|\mathcal{R}y^\delta - x^\dagger\| \mid y^\delta \in Y, \|y - y^\delta\| \leq \delta, x^\dagger \in M_{p,E}\}$$

is called *worst case error* of the method \mathcal{R} on the set $M_{p,E}$. An optimal method \mathcal{R}_{opt} is characterized by $\Delta(\delta, \mathcal{R}_{\text{opt}}) = \inf_{\mathcal{R}} \Delta(\delta, \mathcal{R})$, and this quantity is called *best possible worst case error* on the set $M_{p,E}$. Under Assumption A1(ii) the best possible worst case error can be estimated from below by

$$\inf_{\mathcal{R}} \Delta(\delta, \mathcal{R}) \geq E \left[\psi_p^{-1} \left(\frac{\delta}{ME} \right) \right]^p \quad \text{with} \quad \psi_p(t) := t^p \varrho(t) \quad (2.11)$$

provided $\delta/(ME)$ is an element of the spectrum of the operator $\psi_p(G)$, see [37] and [25, proof of Theorem 2.2]. This lower bound will serve us as benchmark for the best possible accuracy for identifying x^\dagger from noisy data $y^\delta \in Y$ under the link assumption A1 and the smoothness assumption A2.

2.3. Exploiting operator monotonicity. Operator monotonicity has been applied before in different papers for deriving order optimal error bounds in regularization under general smoothness assumptions, see, e. g., [3, 21, 27]. In this section we are going to derive some consequences of Assumption A1 by using operator monotonicity of certain functions. Considering the error representations (2.8) – (2.10) we see that two types of estimates are helpful for deriving error bounds for the noise amplification error and the regularization error in the $\|Ax\|$ – norm and the $\|x\|_p$ – norm, namely estimates of the type

$$\|f_1(T^*T)G^{-s}x\| \leq \|G^{-p}x\| \quad \text{and} \quad \|G^{s-p}x\| \leq \|f_2(T^*T)x\|$$

with certain functions f_1, f_2 and some constant $p > 0$ from Assumption A2. We will derive such estimates from Assumption A1 by using the concept of operator monotone functions which is based on the concept of semiordering. Note that for two nonnegative, self-adjoint bounded linear operators $S_1, S_2 \in \mathcal{L}(X)$ the *semiordering* $S_1 \leq S_2$ is defined by $(S_1x, x) \leq (S_2x, x)$ for all $x \in X$, or equivalently, by $\|S_1^{1/2}x\| \leq \|S_2^{1/2}x\|$ for all $x \in X$.

Definition 2.3. An index function $f : (0, a] \rightarrow \mathbb{R}$ is called *operator monotone* if and only if for any pair of self-adjoint linear operators S_1, S_2 with spectra in $(0, a]$, the relation $S_1 \leq S_2$ implies the relation $f(S_1) \leq f(S_2)$.

Properties and examples for operator monotone functions may be found in [2, 20, 38]. Our further study is based on several functions. The first function is

$$\psi_r(\lambda) = \lambda^r \varrho(\lambda), \quad \psi_r : (0, a] \rightarrow (0, a^r \varrho(a)] \quad (2.12)$$

with ϱ from Assumption A1, $a = \|G\|$ and arbitrary constant r for which $\lambda^r \varrho(\lambda)$ is monotonically increasing. Two other functions h and w are

$$h(t) = \left[\psi_s^{-1}(\sqrt{t}) \right]^{p-s}, \quad w(t) = 1/h(t) \quad (2.13)$$

with constant s from (2.1) and p from Assumption A2.

Remark 2.4. The function h defined by (2.13) possesses the following properties:

- (i) Due to the identity $\sqrt{t}h(t) = \psi_p((\psi_s^2)^{-1}(t))$, the function $t \rightarrow \sqrt{t}h(t)$ is an index function and hence monotonically increasing in the both cases $s \geq p$ and $s \leq p$.

- (ii) In the case of high order regularization with $s \geq p$, the function h is non-increasing. Hence, for $s \geq p$, the function $t \rightarrow h(t)/\sqrt{t}$ is always monotonically decreasing.

In our next proposition we derive some estimates by using Assumption A1.

Proposition 2.5. *Let h and w be defined by (2.13). Then,*

$$\|w(T^*T/M^2)G^{-s}x\| \leq \|G^{-p}x\| \quad \text{under A1(ii) and } w^2 \text{ operator monotone,} \quad (2.14)$$

$$\|w(T^*T/m^2)G^{-s}x\| \leq \|G^{-p}x\| \quad \text{under A1(i) and } h^2 \text{ operator monotone,} \quad (2.15)$$

$$\|G^{s-p}x\| \leq \|w(T^*T/m^2)x\| \quad \text{under A1(i) and } w^2 \text{ operator monotone,} \quad (2.16)$$

$$\|G^{s-p}x\| \leq \|w(T^*T/M^2)x\| \quad \text{under A1(ii) and } h^2 \text{ operator monotone.} \quad (2.17)$$

Proof. First, let us prove (2.14). It follows from Assumption A1(ii) that

$$\|AG^s x\| = \|Tx\| = \|(T^*T)^{1/2}x\| \leq M\|\varrho(G)G^s x\|,$$

which may be written in the equivalent form $T^*T/M^2 \leq \varrho^2(G)G^{2s}$. By using the function ψ_s defined by (2.12), this estimate can be written as $T^*T/M^2 \leq \psi_s^2(G)$. Since $w^2 := 1/h^2$ is assumed to be operator monotone and since $w^2(\psi_s^2(G)) = G^{2s-2p}$ we obtain that $w^2(T^*T/M^2) \leq G^{2s-2p}$ which gives $\|w(T^*T/M^2)x\| \leq \|G^{s-p}x\|$ and hence (2.14). Second, let us prove (2.15). The link condition A1(i) may be written as $\psi_s^2(G) \leq T^*T/m^2$. Since h^2 is assumed to be operator monotone and since $h^2(\psi_s^2(G)) = G^{2p-2s}$ we obtain that $G^{2p-2s} \leq h^2(T^*T/m^2)$. Since $t \rightarrow -1/t$ is operator monotone there follows that $w^2(T^*T/m^2) \leq G^{2s-2p}$, which gives (2.15). The proof of the estimates (2.16) and (2.17) is similar. \square

Example 2.6. (*Finitely smoothing case*). Let us assume that the operators A^*A and G are related by

$$A^*A = G^{2a} \quad (2.18)$$

where a is some positive constant. In this case both Assumptions A1(i) and A1(ii) hold true as equality with $\varrho(\lambda) = \lambda^a$, $m = 1$ and $M = 1$. We easily see that the function ϱ is an index function and that the function ψ_s defined in (2.12) attains the form $\psi_s(\lambda) = \lambda^{a+s}$. Since $\psi_s^{-1}(\sqrt{t}) = t^{1/(2a+2s)}$ we obtain that the functions h and w defined in (2.13) possess the representations

$$h(t) = t^{\frac{p-s}{2(a+s)}}, \quad w(t) = t^{\frac{s-p}{2(a+s)}}.$$

Power functions t^ν are operator monotone for $0 \leq \nu \leq 1$, see [20]. Hence, under the natural side conditions $p \geq 0$, $a > 0$ and $s > -a$ we obtain

- (i) that w^2 is an operator monotone function for $s \geq p$,
- (ii) that h^2 is an operator monotone function for $s \leq p \leq 2s + a$.

2.4. Interpolation in variable Hilbert scales. By interpolation in variable Hilbert scales we can estimate the intermediate norm $\|x\|$ if estimates for some weaker norm $\|\varrho(G)x\|$ and some stronger norm $\|x\|_r$ are known. Variable Hilbert scale inequalities have been introduced by Hegland, see [9, 10]. Such inequalities which extend the classical interpolation inequality became a powerful tool

in the analysis of regularization under general smoothness conditions, see, e.g., [18, 19, 25, 26, 32, 37]. Variable Hilbert scale interpolation is sometimes also called interpolation with a function parameter, see [1, 23]. In our paper we are aiming to combine special variable Hilbert scale inequalities with tools from operator monotonicity.

Proposition 2.7. *Assume $r \geq 0$, $\|x\|_r \leq c_1$ and $\|\varrho(G)x\| \leq c_2$ with some index function ϱ and constants c_1, c_2 . Let $\xi_r(t) := \psi_r^2(t^{1/(2r)})$ be convex where ψ_r is given by (2.12). Then,*

$$\|x\| \leq c_1 \left[\psi_r^{-1} \left(\frac{c_2}{c_1} \right) \right]^r. \quad (2.19)$$

Proof. Let E_λ the spectral family of G^{-2r} . Since ξ_r is convex we may employ Jensen's inequality and obtain

$$\xi_r \left(\frac{\|x\|^2}{\|x\|_r^2} \right) \leq \frac{\int \xi_r(\lambda^{-1}) \lambda d\|E_\lambda x\|^2}{\|x\|_r^2} = \frac{\|\varrho(G)x\|^2}{\|x\|_r^2},$$

or equivalently, $\|x\|_r \cdot \psi_r \left(\|x\|^{1/r} / \|x\|_r^{1/r} \right) \leq \|\varrho(G)x\|$. Since $\varrho(t) = t^{-r} \psi_r(t)$ is increasing we obtain that $t \rightarrow t \psi_r(1/t^{1/r})$ is decreasing. Hence,

$$c_1 \cdot \psi_r \left(\frac{\|x\|^{1/r}}{c_1^{1/r}} \right) \leq \|x\|_r \cdot \psi_r \left(\frac{\|x\|^{1/r}}{\|x\|_r^{1/r}} \right) \leq \|\varrho(G)x\| \leq c_2.$$

Rearranging terms gives (2.19). \square

In our next proposition we provide a further estimate which is based on interpolation arguments.

Proposition 2.8. *Let x_n be the regularized solution (2.1) with exact data y , let ϱ be an arbitrary index function, let Assumption A2 hold and assume $0 \leq s \leq p$. Let in addition ψ_s be defined by (2.12) and*

$$f(t) := \psi_s^2(t^{1/(2p-2s)}) \quad (2.20)$$

be convex. Then, for all regularization parameters σ_n ,

$$\|x_n - x^\dagger\|_{2s-p} \leq E \left[\psi_p^{-1} \left(\frac{\|\varrho(G)(x_n - x^\dagger)\|}{E} \right) \right]^{2p-2s}. \quad (2.21)$$

Proof. Let us introduce the abbreviation $z = x^\dagger - x_n$. From (2.6) we have the identity $G^{-s}z = r_n(T^*T)G^{-s}(x^\dagger - x_0)$, and due to $r_n(\lambda) \leq 1$ we have the estimate $\|r_n^{1/2}(T^*T)\| \leq 1$. We use these properties and obtain due to Cauchy Schwarz inequality and Assumption A2 that

$$\begin{aligned} \|z\|_s^2 &= \|r_n(T^*T)G^{-s}(x^\dagger - x_0)\|^2 \\ &\leq \|r_n^{1/2}(T^*T)G^{-s}(x^\dagger - x_0)\|^2 \\ &= (G^{p-2s}z, G^{-p}(x^\dagger - x_0)) \\ &\leq E\|z\|_{2s-p}. \end{aligned} \quad (2.22)$$

In the case $s = p$, (2.21) follows from (2.22). In the case $0 \leq s < p$, our next aim consists in deriving a second estimate that relates the intermediate norm $\|z\|_{2s-p}$ with the weaker norm $\|\varrho(G)z\|$ and the stronger norm $\|z\|_s$. We derive this estimate by interpolation in variable Hilbert scales. Since f is convex we may employ Jensen's inequality and have

$$f\left(\frac{\|z\|_{2s-p}^2}{\|z\|_s^2}\right) = f\left(\frac{\int \lambda^{s-p} \cdot \lambda^s d\|E_\lambda z\|^2}{\int \lambda^s d\|E_\lambda z\|^2}\right) \leq \frac{\int f(\lambda^{s-p}) \cdot \lambda^s d\|E_\lambda z\|^2}{\int \lambda^s d\|E_\lambda z\|^2}$$

where E_λ is the spectral family of G^{-2} . Since $f(\lambda^{s-p})\lambda^s = \varrho^2(\lambda^{-1/2})$ we obtain

$$f\left(\frac{\|z\|_{2s-p}^2}{\|z\|_s^2}\right) \leq \frac{\int \varrho^2(\lambda^{-1/2}) d\|E_\lambda z\|^2}{\|z\|_s^2} = \frac{\|\varrho(G)z\|^2}{\|z\|_s^2}. \quad (2.23)$$

Now, let us eliminate $\|z\|_s$ in estimate (2.23). We write estimate (2.22) in the equivalent form

$$\|z\|_{2s-p}^{1/2}/E^{1/2} \leq \|z\|_{2s-p}/\|z\|_s \quad (2.24)$$

and introduce two auxiliary functions g and r by

$$g(t) := t^{-2}f(t^2) \quad \text{and} \quad r(t) := tf(t) = \psi_p^2(t^{1/(2p-2s)}). \quad (2.25)$$

Since f is convex and $f(0) = 0$, g is monotonically increasing. Hence, by (2.24), the monotonicity of g and (2.23),

$$g\left(\frac{\|z\|_{2s-p}^{1/2}}{E^{1/2}}\right) \leq g\left(\frac{\|z\|_{2s-p}}{\|z\|_s}\right) = \frac{\|z\|_s^2}{\|z\|_{2s-p}^2} f\left(\frac{\|z\|_{2s-p}^2}{\|z\|_s^2}\right) \leq \frac{\|\varrho(G)z\|^2}{\|z\|_{2s-p}^2}.$$

Multiplying by $\|z\|_{2s-p}^2/E^2$ gives

$$r\left(\frac{\|z\|_{2s-p}}{E}\right) \leq \frac{\|\varrho(G)z\|^2}{E^2}.$$

Since the inverse r^{-1} has the form $r^{-1}(\lambda) = [\psi_p^{-1}(\sqrt{\lambda})]^{2p-2s}$, we obtain (2.21). \square

3. A priori parameter choice

In this section we make use of Proposition 2.5 for estimating the total error in different norms in case the regularization parameter σ_n from (1.3) is chosen *a priori* by

$$\sigma_n^{-1} = \frac{\delta^2}{E^2} \left[\psi_p^{-1}\left(\frac{\delta}{mE}\right) \right]^{2(s-p)}. \quad (3.1)$$

We note that in the finitely smoothing case of Example 2.6 the *a priori* parameter choice (3.1) attains the form $\sigma_n^{-1} = m^2 \left(\frac{\delta}{mE}\right)^{2(s+a)/(a+p)}$.

3.1. Error bounds in the $\|Ax\|$ – norm. We start by providing error bounds for arbitrary $\sigma_n > 0$.

Proposition 3.1. *Let x_n^δ be defined by (2.1), h and w be defined by (2.13) and assume the solution smoothness A2.*

- (i) *High order regularization ($s \geq p$): If $w^2 := 1/h^2$ is operator monotone, then under the link condition A1(ii),*

$$\|Ax_n^\delta - Ax^\dagger\| \leq \delta + E\sqrt{\sigma_n^{-1}}h(\sigma_n^{-1}/M^2). \quad (3.2)$$

- (ii) *Low order regularization ($s \leq p$): If h^2 is operator monotone and if $h(t)/\sqrt{t}$ is decreasing, then under the link condition A1(i),*

$$\|Ax_n^\delta - Ax^\dagger\| \leq \delta + E\sqrt{\sigma_n^{-1}}h(\sigma_n^{-1}/m^2). \quad (3.3)$$

Proof. For estimating $\|Ax_n^\delta - Ax_n\|$ we use the error representation (2.7) and obtain due to $\lambda g_n(\lambda) \leq 1$, see (2.3), the estimate

$$\|Ax_n^\delta - Ax_n\| = \|Tg_n(T^*T)T^*(y^\delta - y)\| \leq \delta \sup_{\lambda} |\lambda g_n(\lambda)| \leq \delta. \quad (3.4)$$

For estimating $\|Ax_n - Ax^\dagger\|$ in the high order case $s \geq p$ we use the error representation (2.8), exploit Assumption A2 and estimate (2.14) which requires operator monotonicity of w^2 and the second link condition A1(ii) and obtain

$$\|Ax_n - Ax^\dagger\| = \|Tr_n(T^*T)G^{-s}(x^\dagger - x_0)\| \leq \|Tr_n(T^*T)h(T^*T/M^2)\| \cdot E \quad (3.5)$$

For estimating the norm term in (3.5) we distinguish two cases $\lambda \leq \sigma_n^{-1}$ and $\lambda \geq \sigma_n^{-1}$. In the first case $\lambda \leq \sigma_n^{-1}$ we use $g_n(\lambda) \geq 0$, or equivalently $r_n(\lambda) \leq 1$, exploit the increasing behavior of $\sqrt{t}h(t)$ that holds true due to Remark 2.4 (i) and obtain

$$\sqrt{\lambda}r_n(\lambda)h(\lambda/M^2) \leq \sqrt{\lambda}h(\lambda/M^2) \leq \sqrt{\sigma_n^{-1}}h(\sigma_n^{-1}/M^2).$$

In the second case $\lambda \geq \sigma_n^{-1}$ we use $\lambda r_n(\lambda) \leq \sigma_n^{-1}$, see (2.4), exploit the decreasing behavior of $h(t)/\sqrt{t}$ that holds true due to Remark 2.4 (ii) and obtain

$$\sqrt{\lambda}r_n(\lambda)h(\lambda/M^2) \leq \sigma_n^{-1}h(\lambda/M^2)/\sqrt{\lambda} \leq \sqrt{\sigma_n^{-1}}h(\sigma_n^{-1}/M^2).$$

From the both cases we obtain that (3.5) attains the form

$$\|Ax_n - Ax^\dagger\| \leq E\sqrt{\sigma_n^{-1}}h(\sigma_n^{-1}/M^2).$$

From this estimate and (3.4) we obtain (3.2). For the proof of part (ii) we proceed analogously by exploiting (2.15) instead of (2.14). \square

Remark 3.2. Let us discuss the monotonicity condition in part (ii) of Proposition 3.1 for the finitely smoothing case of Example 2.6. For this example we have

$$h(t)/\sqrt{t} = t^{\frac{p-2s-a}{2(a+s)}}.$$

Hence, $h(t)/\sqrt{t}$ is decreasing for $p \leq 2s + a$. This coincides with Natterer's side condition in (1.4).

Corollary 3.3. *Let be satisfied the assumptions of Proposition 3.1 and let σ_n be chosen by (3.1). Then, in the both cases (i) and (ii) of Proposition 3.1 we have*

$$(i) \quad \|Ax_n^\delta - Ax^\dagger\| \leq (1 + M/m) \delta, \quad (3.6)$$

$$(ii) \quad \|Ax_n^\delta - Ax^\dagger\| \leq 2\delta. \quad (3.7)$$

Proof. Let us prove estimate (3.6) for the high order case (i). The *a priori* parameter choice (3.1) can be written in the equivalent form

$$E\sqrt{\sigma_n^{-1}} = \delta w(\sigma_n^{-1}/m^2). \quad (3.8)$$

Hence, by (3.2) and (3.8),

$$\|Ax_n^\delta - Ax^\dagger\| \leq \delta + \delta h(\sigma_n^{-1}/M^2) w(\sigma_n^{-1}/m^2). \quad (3.9)$$

Since $\sigma_n^{-1}/M^2 \leq \sigma_n^{-1}/m^2$ and since $\sqrt{t}h(t)$ is increasing we have

$$h(\sigma_n^{-1}/M^2) \leq \frac{M}{m} h(\sigma_n^{-1}/m^2).$$

From this estimate and (3.9) we obtain (3.6). The proof for the estimate (3.7) for the low order case (ii) is analogous. \square

3.2. Error bounds in X_p . We start by providing error bounds with respect to the $\|\cdot\|_p$ -norm for arbitrary $\sigma_n > 0$.

Proposition 3.4. *Let x_n^δ be defined by (2.1), h and w be defined by (2.13) and assume the link condition A1 and the solution smoothness A2.*

(i) *High order regularization ($s \geq p$): If $w^2 := 1/h^2$ is operator monotone,*

$$\|x_n^\delta - x^\dagger\|_p \leq \delta \sqrt{\sigma_n} w(\sigma_n^{-1}/m^2) + E \cdot M/m. \quad (3.10)$$

(ii) *Low order regularization ($s \leq p$): If h^2 is operator monotone and if $h(t)/\sqrt{t}$ is decreasing,*

$$\|x_n^\delta - x^\dagger\|_p \leq \delta \sqrt{\sigma_n} w(\sigma_n^{-1}/M^2) + E \cdot M/m. \quad (3.11)$$

Proof. Let us consider the high order case (i). For estimating $\|x_n^\delta - x_n\|_p$ we use the error representation (2.9) and obtain due to the estimate (2.16) of Proposition 2.5 the estimate

$$\|x_n^\delta - x_n\|_p = \|G^{s-p} g_n(T^*T) T^*(y^\delta - y)\| \leq \delta \left\| w \left(\frac{1}{m^2} T^*T \right) g_n(T^*T) T^* \right\|. \quad (3.12)$$

For estimating the norm term in (3.12) we distinguish two cases $\lambda \leq \sigma_n^{-1}$ and $\lambda \geq \sigma_n^{-1}$. In the first case $\lambda \leq \sigma_n^{-1}$ we use $g_n(\lambda) \leq \sigma_n$, see (2.2), exploit the increasing behavior of $\sqrt{t}w(t)$ that follows since due to Remark 2.4 (ii) the function $h(t)/\sqrt{t}$ is decreasing and obtain

$$w(\lambda/m^2) g_n(\lambda) \sqrt{\lambda} \leq w(\lambda/m^2) \sqrt{\lambda} \sigma_n \leq w(\sigma_n^{-1}/m^2) \sqrt{\sigma_n}.$$

In the second case $\lambda \geq \sigma_n^{-1}$ we use $\lambda g_n(\lambda) \leq 1$, exploit the decreasing behavior of $w(t)/\sqrt{t}$ that follows since due to Remark 2.4 (i) the function $\sqrt{t}h(t)$ is increasing and obtain

$$w(\lambda/m^2) g_n(\lambda) \sqrt{\lambda} \leq w(\lambda/m^2) / \sqrt{\lambda} \leq w(\sigma_n^{-1}/m^2) \sqrt{\sigma_n}.$$

From the both cases we obtain that (3.12) attains the form

$$\|x_n^\delta - x_n\|_p \leq \delta \sqrt{\sigma_n} w(\sigma_n^{-1}/m^2). \quad (3.13)$$

For estimating $\|x_n - x^\dagger\|_p$ in the high order case (i) we use the error representation (2.10), exploit the estimate (2.14) of Proposition 2.5, use in addition the estimate (2.16) and obtain due to $\|x^\dagger - x_0\|_p \leq E$ and $g_n(\lambda) \geq 0$, or equivalently $r_n(\lambda) \leq 1$, the estimate

$$\begin{aligned} \|x^\dagger - x_n\|_p &= \|G^{s-p} r_n(T^*T) G^{-s} (x^\dagger - x_0)\| \\ &\leq \|w(T^*T/m^2) r_n(T^*T) G^{-s} (x^\dagger - x_0)\| \\ &\leq E \|w(T^*T/m^2) r_n(T^*T) h(T^*T/M^2)\| \\ &\leq E \sup_{\lambda} |h(\lambda/M^2) w(\lambda/m^2)|. \end{aligned} \quad (3.14)$$

Due to Remark 2.4 (i) the function $\sqrt{t}h(t)$ is increasing, or equivalently, $w(t)/\sqrt{t}$ is decreasing. Hence, from $\lambda/M^2 \leq \lambda/m^2$ we have $w(\lambda/m^2) \leq \frac{M}{m} w(\lambda/M^2)$, and (3.14) attains the form

$$\|x^\dagger - x_n\|_p \leq E \cdot M/m. \quad (3.15)$$

From (3.13) and (3.15) we obtain (3.10). For the proof of part (ii) we proceed analogously by exploiting (2.15) and (2.17) instead of (2.14) and (2.16). \square

From Proposition 3.4 we have along the line of Corollary 3.3 the following

Corollary 3.5. *Let be satisfied the assumptions of Proposition 3.4 and let σ_n be chosen by (3.1). Then, in the both cases (i) and (ii) of Proposition 3.4 we have*

$$(i) \quad \|x_n^\delta - x^\dagger\|_p \leq E \cdot (1 + M/m), \quad (3.16)$$

$$(ii) \quad \|x_n^\delta - x^\dagger\|_p \leq 2E \cdot M/m. \quad (3.17)$$

3.3. Error bounds in X . For deriving order optimal error bounds for the total error $\|x_n^\delta - x^\dagger\|$ with σ_n chosen a priori by (3.1) we employ interpolation techniques from Proposition 2.7 and use the results of Corollary 3.5 which provides a bound for $\|x_n^\delta - x^\dagger\|_p$ and the results of Corollary 3.3 which together with the first link condition A1(i) provides a bound for $\|\rho(G)(x_n^\delta - x^\dagger)\|$.

Theorem 3.6. *Let be satisfied the assumptions of Proposition 3.1 and 3.4 and let σ_n be chosen a priori by (3.1). If the function $\xi_p(t) := \psi_p^2(t^{1/(2p)})$ is convex, then x_n^δ is order optimal on the set $M_{p,E}$ in the both cases of high order regularization $s \geq p$ and low order regularization $s \leq p$. In fact, in both cases,*

$$\|x_n^\delta - x^\dagger\| \leq c_1 [\psi_p^{-1}(c_2 \delta)]^p \quad (3.18)$$

with some constants c_1 and c_2 which can be extracted from the proof.

Proof. Due to Corollary 3.5, Corollary 3.3 and Assumption A1(i), in both cases of high- and low order regularization we have

$$\|x_n^\delta - x^\dagger\|_p \leq k_1 \quad \text{and} \quad \|\varrho(G)(x_n^\delta - x^\dagger)\| \leq k_2 \delta$$

with some constants k_1 and k_2 . Using the interpolation estimate (2.19) of Proposition 2.7 yields (3.18). \square

Note that for the finitely smoothing case of Example 2.6 we have $\xi_p(t) = t^{(a+p)/p}$, which is convex for arbitrary $p > 0$.

3.4. Revisiting the low order case. The error bounds given in Subsection 3.3 require in both cases of high order and low order regularization the both link conditions A1(i) and A1(ii). We will show in this subsection that in the case of low order regularization $s \leq p$ order optimal error bounds can be obtained without the second link condition A1(ii). However, this will only be possible for $s \geq 0$. We exploit in our study the property

$$\sqrt{\lambda}g_n(\lambda) \leq \sqrt{\sigma_n} \quad \text{for } \lambda \in (0, \|T\|^2] \quad (3.19)$$

which follows from the both properties (2.2) and (2.3) of Proposition 2.1 and start by providing some error bound in the $\|\cdot\|_s$ -norm for arbitrary $\sigma_n > 0$.

Proposition 3.7. *Let x_n^δ be defined by (2.1), h be defined by (2.13) and assume the solution smoothness A2. If h^2 is operator monotone and $h(t)/t$ is decreasing, then under the link condition A1(i),*

$$\|x_n^\delta - x^\dagger\|_s \leq \delta \sqrt{\sigma_n} + Eh(\sigma_n^{-1}/m^2). \quad (3.20)$$

Proof. For estimating $\|x_n^\delta - x_n\|_s$ in the low order case $s \leq p$ we use the error representation (2.6) and obtain due to $\sqrt{\lambda}g_n(\lambda) \leq \sqrt{\sigma_n}$, see (3.19), the estimate

$$\|x_n^\delta - x_n\|_s = \|g_n(T^*T)T^*(y^\delta - y)\| \leq \delta \sqrt{\sigma_n}. \quad (3.21)$$

For estimating $\|x_n - x^\dagger\|_s$ we use the error representation (2.6), exploit the estimate (2.15) of Proposition 2.5 and obtain due to $\|x^\dagger - x_0\|_p \leq E$ the estimate

$$\|x^\dagger - x_n\|_s = \|r_n(T^*T)G^{-s}(x^\dagger - x_0)\| \leq E \|r_n(T^*T)h(T^*T/m^2)\|. \quad (3.22)$$

For estimating the norm term in (3.22) we distinguish two cases $\lambda \leq \sigma_n^{-1}$ and $\lambda \geq \sigma_n^{-1}$. In the first case $\lambda \leq \sigma_n^{-1}$ we use $r_n(\lambda) \leq 1$, or equivalently, $g_n(\lambda) \geq 0$ exploit the increasing behavior of $h(t)$ which is always satisfied since h^2 is operator monotone and obtain

$$r_n(\lambda)h(\lambda/m^2) \leq h(\lambda/m^2) \leq h(\sigma_n^{-1}/m^2).$$

In the second case $\lambda \geq \sigma_n^{-1}$ we use $\lambda r_n(\lambda) \leq \sigma_n^{-1}$, exploit the decreasing behavior of $h(t)/t$ and obtain

$$r_n(\lambda)h(\lambda/m^2) \leq \sigma_n^{-1}h(\lambda/m^2)/\lambda \leq h(\sigma_n^{-1}/m^2).$$

From the both cases we obtain that (3.22) attains the form $\|x^\dagger - x_n\|_s \leq Eh(\sigma_n^{-1}/m^2)$. From this estimate and (3.21) we obtain (3.20). \square

Since the parameter choice (3.1) can be written in the equivalent form $\delta\sqrt{\sigma_n} = Eh(\sigma_n^{-1}/m^2)$ we obtain from Proposition 3.7 the following

Corollary 3.8. *Let be satisfied the assumptions of Proposition 3.7 and let σ_n be chosen a priori by (3.1). Then,*

$$\|x_n^\delta - x^\dagger\|_s \leq 2E \left[\psi_p^{-1} \left(\frac{\delta}{mE} \right) \right]^{p-s}. \quad (3.23)$$

Proof. From (2.12) we have $\varrho(\lambda) = \lambda^{-s}\psi_s(\lambda)$ and $\varrho(\lambda) = \lambda^{-p}\psi_p(\lambda)$. Consequently,

$$\psi_s(\lambda) = \lambda^{s-p}\psi_p(\lambda) \quad \text{and} \quad \psi_p(\lambda) = \lambda^{p-s}\psi_s(\lambda). \quad (3.24)$$

We use the first equation of (3.24), substitute $\lambda = \psi_p^{-1}(t)$ and obtain $\psi_s(\psi_p^{-1}(t)) = t[\psi_p^{-1}(t)]^{s-p}$. From this equation we conclude that the *a priori* parameter choice (3.1), which is equivalent to $\sigma_n^{-1/2}/m = \frac{\delta}{mE} [\psi_p^{-1}(\frac{\delta}{mE})]^{s-p}$, can be rewritten as $\sigma_n^{-1/2}/m = \psi_s(\psi_p^{-1}(\frac{\delta}{mE}))$, or equivalently,

$$\psi_s^{-1}\left(\frac{\sigma_n^{-1/2}}{m}\right) = \psi_p^{-1}\left(\frac{\delta}{mE}\right). \quad (3.25)$$

Clearly, (3.25) is equivalent to $\psi_p\left(\psi_s^{-1}\left(\sigma_n^{-1/2}/m\right)\right) = \frac{\delta}{mE}$. We use the second equation of (3.24) and write this equation in the form

$$\left[\psi_s^{-1}\left(\frac{\sigma_n^{-1/2}}{m}\right)\right]^{p-s} \frac{\sigma_n^{-1/2}}{m} = \frac{\delta}{mE}.$$

From this equation and the definition of h by (2.13) we see that the parameter choice (3.1) can be written in the equivalent form $\delta\sqrt{\sigma_n} = Eh(\sigma_n^{-1}/m^2)$. Due to this equation, estimate (3.20) of Proposition 3.7 attains the form

$$\|x_n^\delta - x^\dagger\|_s \leq 2\delta\sqrt{\sigma_n}. \quad (3.26)$$

From this estimate and the parameter choice (3.1) we obtain (3.23). \square

Now, by using the both estimates (3.23) and (3.7), we obtain the following order optimality result for x_n^δ on the set $M_{p,E}$.

Theorem 3.9. *Let x_n^δ be defined by (2.1) with σ_n chosen by (3.1), assume the link condition A1(i) and the solution smoothness A2. If h^2 is operator monotone, $h(t)/\sqrt{t}$ is decreasing and $\xi_s(t) := \psi_s^2(t^{1/(2s)})$ is convex, then*

$$\|x_n^\delta - x^\dagger\| \leq 2E \left[\psi_p^{-1}\left(\frac{\delta}{mE}\right)\right]^p. \quad (3.27)$$

Proof. From estimate (3.7) and Assumption A1(i) we have the estimate

$$\|\varrho(G)(x_n^\delta - x^\dagger)\| \leq 2\delta/m. \quad (3.28)$$

We apply the interpolation estimate (2.19) of Proposition 2.7 and obtain together with (3.26) and (3.28) the estimate

$$\|x_n^\delta - x^\dagger\| \leq 2\delta\sigma_n^{1/2} [\psi_s^{-1}(\sigma_n^{-1/2}/m)]^s. \quad (3.29)$$

It remains to show that for the parameter choice (3.1) both right hand sides of (3.27) and (3.29) coincide. We use formula (3.25) and obtain that (3.29) can be written in the equivalent form

$$\|x_n^\delta - x^\dagger\| \leq 2\delta\sigma_n^{1/2} \left[\psi_p^{-1}\left(\frac{\delta}{mE}\right)\right]^s. \quad (3.30)$$

Now we rewrite (3.1) as $\sigma_n^{1/2} = \frac{E}{\delta} [\psi_p^{-1}(\frac{\delta}{mE})]^{p-s}$, substitute this into (3.30) and obtain (3.27). \square

4. Discrepancy principle

If the constants m and E in the *a priori* parameter choice (3.1) are unknown, then the parameter choice $\sigma_n^{-1} = \left(\frac{\delta}{c_2}\right)^2 \left[\psi_p^{-1}\left(\frac{\delta}{c_1 c_2}\right)\right]^{2(s-p)}$ may be used where c_1 and c_2 are positive constants guessing m and E , respectively. For this parameter choice, the order optimality results of Theorems 3.6 and 3.9 still hold true. In case of rough estimates for m and E , and in particular in cases where p and ψ_p are unknown, *a posteriori* rules for choosing σ_n have to be used. In the discrepancy principle (see [24]) the regularization parameter σ_n is chosen as the solution of the nonlinear equation

$$d(\sigma_n) := \|Ax_n^\delta - y^\delta\| = C\delta \quad (4.1)$$

with some constant $C \geq 1$. For practical reasons it makes sense to choose σ_n such that

$$C_1\delta \leq d(\sigma_n) \leq C_2\delta \quad (4.2)$$

with some constants C_1, C_2 that obey $1 \leq C_1 \leq C_2$. In computations it makes sense to choose C_2 with $C_2 > C_1$.

Remark 4.1. For realizing the discrepancy principle (4.1) or (4.2) approximately, one practical way is as follows. We start with some large α_1 in (1.2), use a decreasing α -sequence and iterate as long as the discrepancy is in the magnitude of the noise level. More accurately, we consider the decreasing sequence $\Delta = \{\alpha_k\}_{k=1}^\infty$ and choose n as the first integer for which

$$\|Ax_n^\delta - y^\delta\| \leq C\delta < \|Ax_k^\delta - y^\delta\|, \quad 0 \leq k < n \quad (4.3)$$

with some $C > 1$. Some care is required for the final iteration step in which one has to take care that the discrepancy becomes not too small and remains in the magnitude of δ . This can be guaranteed by assuming that the final α_n is not too small and obeys

$$1/\alpha_n \leq c\sigma_{n-1} \quad (4.4)$$

with some positive constant c . For the geometric sequence $\Delta = \{q^{k-1}\alpha_1\}_{k=1}^\infty$ with some $q < 1$, assumption (4.4) is satisfied with $c = 1/q$, see [7]. We show in Subsection 4.3 that for the version (4.3) of the discrepancy principle, analogous convergence rate results to that of the *a posteriori* rule (4.2) hold true.

4.1. Properties. Due to (2.1), the discrepancy $y^\delta - Ax_n^\delta$ can be represented by

$$y^\delta - Ax_n^\delta = r_n(TT^*)(y^\delta - Ax_0) = \left(\prod_{k=1}^n \alpha_k(TT^* + \alpha_k I)^{-1}\right)(y^\delta - Ax_0). \quad (4.5)$$

From this representation we conclude that the discrepancy is monotonically decreasing with respect to the iteration number, that is,

$$\|y^\delta - Ax_k^\delta\| < \|y^\delta - Ax_{k-1}^\delta\|, \quad k = 1, 2, \dots$$

For $\sigma_n \rightarrow \infty$ we have $r_n(\lambda) \rightarrow 0$, and for $\sigma_n \rightarrow 0$ we have $r_n(\lambda) \rightarrow 1$. Therefore, by (4.5), we have the two limit relations

$$\lim_{\sigma_n \rightarrow \infty} \|y^\delta - Ax_n^\delta\| = 0 \quad \text{and} \quad \lim_{\sigma_n \rightarrow 0} \|y^\delta - Ax_n^\delta\| = \|y^\delta - Ax_0\|.$$

From both limit relations we conclude that under the condition $\|y^\delta - Ax_0\| > C\delta$ there exists σ_n (not necessarily unique) that obeys rule (4.1) or rule (4.3), respectively, and that under the condition $\|y^\delta - Ax_0\| > C_2\delta$ there exists σ_n that obeys rule (4.2).

Now we assume that for some given σ_{n-1} we have $\|Ax_{n-1}^\delta - y^\delta\| > C\delta$. Then, the discrepancy $d(\alpha_n) := \|Ax_n^\delta - y^\delta\|$ as a function of α_n possesses following properties:

- (i) For $\alpha_n \rightarrow 0$ we have the limit relation $\lim_{\alpha_n \rightarrow 0} d(\alpha_n) = 0$.
- (ii) For $\alpha_n \rightarrow \infty$ we have $\lim_{\alpha_n \rightarrow \infty} d(\alpha_n) = \|Ax_{n-1}^\delta - y^\delta\| > C\delta$.
- (iii) The function $d(\alpha_n)$ is continuous and strictly monotonically increasing.

As a consequence, there exists α_n^* with $d(\alpha_n^*) = C\delta$.

Following proposition gives us some monotonicity property for the error $\|x_n^\delta - x^\dagger\|_s$ with respect to the X_s -norm which tells us that the iteration (1.2) should not be stopped as long as $\|Ax_n^\delta - y^\delta\| \geq \delta$ holds. In the special case $s = 0$, such monotonicity property may be found in [5].

Proposition 4.2. *Let $x^\dagger \in X_s$, let x_n^δ be defined by the iteration (1.2) and let $\|Ax_n^\delta - y^\delta\| \geq \delta$. Then,*

$$\|x_n^\delta - x^\dagger\|_s < \|x_{n-1}^\delta - x^\dagger\|_s. \quad (4.6)$$

Proof. The iteration (1.2) can be rewritten as

$$x_n^\delta = x_{n-1}^\delta + B^{-2s} A^* z_{n-1} \quad \text{with} \quad z_{n-1} = (TT^* + \alpha_n I)^{-1} (y^\delta - Ax_{n-1}^\delta).$$

Consequently, for $d := \|x_n^\delta - x^\dagger\|_s^2 - \|x_{n-1}^\delta - x^\dagger\|_s^2$ we have

$$\begin{aligned} d &= \|x_{n-1}^\delta + B^{-2s} A^* z_{n-1} - x^\dagger\|_s^2 - \|x_{n-1}^\delta - x^\dagger\|_s^2 \\ &= (2x_{n-1}^\delta - 2x^\dagger + B^{-2s} A^* z_{n-1}, B^{-2s} A^* z_{n-1})_s \\ &= (x_{n-1}^\delta + x_n^\delta - 2x^\dagger, B^{-2s} A^* z_{n-1})_s \\ &= (Ax_{n-1}^\delta + Ax_n^\delta - 2y, z_{n-1}) \\ &= (2(y^\delta - y) + (Ax_{n-1}^\delta - y^\delta) + (Ax_n^\delta - y^\delta), z_{n-1}) \\ &\leq 2\|z_{n-1}\| \left(\delta - \frac{((y^\delta - Ax_{n-1}^\delta) + (y^\delta - Ax_n^\delta), z_{n-1})}{2\|z_{n-1}\|} \right). \end{aligned}$$

Let $r_n := y^\delta - Ax_n^\delta$. Then, from (4.5) we have $r_n = \alpha_n (TT^* + \alpha_n I)^{-1} r_{n-1}$. Hence, the element z_{n-1} can be written as $z_{n-1} = \alpha_n^{-1} r_n$. Consequently,

$$\|x_n^\delta - x^\dagger\|_s^2 - \|x_{n-1}^\delta - x^\dagger\|_s^2 \leq \frac{2\|r_n\|}{\alpha_n} \left(\delta - \frac{(r_{n-1} + r_n, r_n)}{2\|r_n\|} \right). \quad (4.7)$$

We use again the identity $r_n = \alpha_n (TT^* + \alpha_n I)^{-1} r_{n-1}$, or equivalently, $r_{n-1} = \alpha_n^{-1} (TT^* + \alpha_n I) r_n$, multiply by r_n and obtain

$$(r_{n-1}, r_n) = \alpha_n^{-1} \|T^* r_n\|^2 + \|r_n\|^2 > \|r_n\|^2. \quad (4.8)$$

From (4.7), (4.8) and $\|r_n\| \geq \delta$ we obtain $\|x_n^\delta - x^\dagger\|_s^2 - \|x_{n-1}^\delta - x^\dagger\|_s^2 < 0$. \square

4.2. Error bounds in X . In this subsection we show that for σ_n chosen by (4.1) or (4.2), respectively, the order optimal error bound (3.18) holds true under analogous assumptions of Theorem 3.6. In a first proposition we provide some estimate for the regularization parameter σ_n chosen by (4.2).

Proposition 4.3. *Let x_n^δ be defined by (2.1), h and w be defined by (2.13), σ_n be chosen by the discrepancy principle (4.2) with $1 < C_1 \leq C_2$ and assume the solution smoothness A2.*

- (i) *High order regularization ($s \geq p$): If $w^2 := 1/h^2$ is operator monotone, then under the link condition A1(ii),*

$$(C_1 - 1)\delta \leq E\sqrt{\sigma_n^{-1}}h(\sigma_n^{-1}/M^2). \quad (4.9)$$

- (ii) *Low order regularization ($s \leq p$): If h^2 is operator monotone and if $h(t)/\sqrt{t}$ is decreasing, then under the link condition A1(i),*

$$(C_1 - 1)\delta \leq E\sqrt{\sigma_n^{-1}}h(\sigma_n^{-1}/m^2). \quad (4.10)$$

Proof. Let us prove part (i). From (2.1) we have $y^\delta - Ax_n^\delta = r_n(TT^*)(y^\delta - Ax_0)$. Due to rule (4.2), the identity $y - Ax_n = r_n(TT^*)(y - Ax_0)$ and the estimate $\|r_n(TT^*)\| \leq 1$ we obtain that

$$\begin{aligned} C_1\delta &\leq \|r_n(TT^*)(y^\delta - Ax_0)\| \\ &\leq \|r_n(TT^*)(y - Ax_0)\| + \|r_n(TT^*)(y - y^\delta)\| \\ &\leq \|y - Ax_n\| + \delta. \end{aligned} \quad (4.11)$$

From the proof of Proposition 3.1 we have that

$$\|y - Ax_n\| \leq E\sqrt{\sigma_n^{-1}}h(\sigma_n^{-1}/M^2).$$

This estimate and (4.11) provide the desired estimate (4.9) and the proof of part (i) is complete. For the proof of part (ii) we proceed in an analogous way, but use instead of (2.14) the estimate (2.15) which requires the link condition A1(i) and the operator monotonicity of the function h^2 . \square

From Propositions 3.4 and 4.3 we obtain that the total error $x_n^\delta - x^\dagger$ is bounded in the $\|\cdot\|_p$ -norm for the *a posteriori* parameter choice σ_n chosen by the discrepancy principle (4.2).

Proposition 4.4. *Let x_n^δ be defined by (2.1), σ_n be chosen by the discrepancy principle (4.2) with $1 < C_1 \leq C_2$ and assume the solution smoothness A2 and the link condition A1.*

- (i) *High order regularization ($s \geq p$): If $w^2 := 1/h^2$ is operator monotone,*

$$\|x_n^\delta - x^\dagger\|_p \leq \frac{E}{C_1 - 1} \cdot \frac{M}{m} + E \cdot \frac{M}{m}. \quad (4.12)$$

- (ii) *Low order regularization ($s \leq p$): If h^2 is operator monotone and $h(t)/\sqrt{t}$ is decreasing, then,*

$$\|x_n^\delta - x^\dagger\|_p \leq \frac{E}{C_1 - 1} \cdot \frac{M}{m} + E \cdot \frac{M}{m}. \quad (4.13)$$

Proof. In the case (i) we exploit the increasing behavior of $\sqrt{t}h(t)$ and conclude from $\sigma_n^{-1}/M^2 \leq \sigma_n^{-1}/m^2$ that $h(\sigma_n^{-1}/M^2) \leq \frac{M}{m}h(\sigma_n^{-1}/m^2)$, which together with part (i) of the two Propositions 3.4 and 4.3 provides (4.12). In the case (ii) we exploit the decreasing behavior of $h(t)/\sqrt{t}$, or equivalently the increasing behavior of $\sqrt{t}w(t)$ and conclude from $\sigma_n^{-1}/M^2 \leq \sigma_n^{-1}/m^2$ that $w(\sigma_n^{-1}/M^2) \leq \frac{M}{m}w(\sigma_n^{-1}/m^2)$, which together with part (ii) of the two Propositions 3.4 and 4.3 provides (4.13). \square

Now we are in a position to prove the main result of this section. In our next theorem we will see that order optimal error bounds can be guaranteed in case σ_n is chosen by the discrepancy principle (4.2) with $1 < C_1 \leq C_2$.

Theorem 4.5. *Let be satisfied the assumptions of Proposition 4.4 and assume in addition that $\xi_p(t) := \psi_p^2(t^{1/(2p)})$ is convex. Then,*

$$\|x_n^\delta - x^\dagger\| \leq c_1 [\psi_p^{-1}(c_2\delta)]^p \quad (4.14)$$

with some constants c_1, c_2 which can be extracted from the proof.

Proof. Due to Proposition 4.4, in both cases (i) and (ii) of high- and low order regularization the total error obeys

$$\|x_n^\delta - x^\dagger\|_p \leq cE \quad (4.15)$$

with some $c \geq 1$ and σ_n chosen by the discrepancy principle (4.2) with $1 < C_1 \leq C_2$. From (4.2) and the triangle inequality we have

$$\|Ax_n^\delta - Ax^\dagger\| \leq \|Ax_n^\delta - y^\delta\| + \|y - y^\delta\| \leq (C_2 + 1)\delta.$$

Using in addition the link condition A1(i) yields

$$\|\rho(G)(x_n^\delta - x^\dagger)\| \leq (C_2 + 1)\delta/m. \quad (4.16)$$

Now the result of the theorem follows from (4.15), (4.16) and Proposition 2.7. \square

4.3. Error bounds for rule (4.3). For the *a posteriori* rule (4.3) of choosing the regularization parametr σ_n , analogous order optimal error bounds to that of Theorem 4.5 can be obtained.

Theorem 4.6. *Let x_n^δ be defined by (2.1), let σ_n be chosen by rule (4.3) where α_n obeys (4.4), let the both Assumptions A1 and A2 hold and assume that $\xi_p(t) := \psi_p^2(t^{1/(2p)})$ is convex. Assume further*

- (i) *in case of high order regularization ($s \geq p$) that $w^2 := 1/h^2$ is operator monotone and*
- (ii) *in case of low order regularization ($s \leq p$) that h^2 is operator monotone and $h(t)/\sqrt{t}$ is decreasing.*

Then, the regularized solution x_n^δ obeys the order optimal error bound

$$\|x_n^\delta - x^\dagger\| \leq c_1 [\psi_p^{-1}(c_2\delta)]^p \quad (4.17)$$

with some constants c_1, c_2 which can be extracted from the proof.

Proof. We give the proof for the high order case $s \geq p$, the proof for the low order case $s \leq p$ is similar. In the *first step* of our proof we proceed according to the proof of Proposition 4.3, exploit that $C\delta \leq \|r_{n-1}(TT^*)(y^\delta - Ax_0)\|$ and obtain

$$(C-1)\delta \leq E\sqrt{\sigma_{n-1}^{-1}}h(\sigma_{n-1}^{-1}/M^2).$$

From this estimate and (3.10) we obtain

$$\|x_n^\delta - x^\dagger\|_p \leq \frac{E}{C-1} \cdot \frac{\sqrt{\sigma_{n-1}^{-1}}h(\sigma_{n-1}^{-1}/M^2)}{\sqrt{\sigma_n^{-1}}h(\sigma_n^{-1}/m^2)} + E \cdot M/m. \quad (4.18)$$

Now we consider two cases. In the *first case* with $\sigma_{n-1}^{-1}/M^2 \leq \sigma_n^{-1}/m^2$ we use the increasing behavior of $\sqrt{t}h(t)$ and obtain from (4.18) the estimate

$$\|x_n^\delta - x^\dagger\|_p \leq \frac{E}{C-1} \cdot \frac{M}{m} + E \cdot M/m.$$

In the *second case* with $\sigma_{n-1}^{-1}/M^2 \geq \sigma_n^{-1}/m^2$ we use the decreasing behavior of h , exploit in addition that due to (4.4) we have $\sigma_n = 1/\alpha_n + \sigma_{n-1} \leq (c+1)\sigma_{n-1}$, or equivalently, $\sigma_{n-1}^{-1} \leq (c+1)\sigma_n^{-1}$, and obtain from (4.18) the estimate

$$\|x_n^\delta - x^\dagger\|_p \leq \frac{E}{C-1} \cdot \sqrt{c+1} + E \cdot M/m.$$

From the both cases we have that $\|x_n^\delta - x^\dagger\|_p$ can be estimated by

$$\|x_n^\delta - x^\dagger\|_p \leq \frac{E}{C-1} \cdot \max\left\{M/m, \sqrt{c+1}\right\} + E \cdot M/m. \quad (4.19)$$

In the *second step* we proceed according to the proof of (4.16) and obtain

$$\|\varrho(G)(x_n^\delta - x^\dagger)\| \leq (C+1)\delta/m. \quad (4.20)$$

In the final *third step* of the proof we use the both estimates (4.19) and (4.20), apply Proposition 2.7 and obtain (4.17). \square

4.4. Discrepancy principle revisited. The error bounds given in Subsection 4.2 require in both cases of high order and low order regularization the both link conditions A1(i) and A1(ii), and the assumption $C_1 > 1$ in the discrepancy principle (4.2). We will show in this subsection that in the case of low order regularization $s \leq p$ order optimal error bounds can be obtained without the second link condition A1(ii). Our estimate in Theorem 4.8 shows that $C_1 = C_2 = 1$ in the discrepancy principle (4.2) is best possible in the sense of minimal error bounds. We start our study with some important inequality.

Proposition 4.7. *For $0 \leq s \leq p$, the regularized solution x_n^δ defined by (2.1) obeys the estimate*

$$\begin{aligned} \|Ax_n^\delta - y^\delta\|^2 + \sigma_n^{-1}\|x_n^\delta - x^\dagger\|_s^2 &\leq \sigma_n^{-1}\|r_n^{1/2}(T^*T)G^{-s}(x^\dagger - x_0)\|^2 \\ &\quad + \|y - y^\delta\|^2. \end{aligned} \quad (4.21)$$

Proof. Let $A : X_s \rightarrow Y$ be the restriction of A to $X_s \subset X$ and $A_s^* : Y \rightarrow X_s$ its adjoint. Due to the valid identity $(Ax, y) = (x, A_s^*y)_s = (x, A^*y) = (x, G^{2s}A^*y)_s$ for all $x \in X_s$ and $y \in Y$ we conclude that the adjoint $A_s^* : Y \rightarrow X_s$ of the operator $A : X_s \rightarrow Y$ is given by $A_s^* = G^{2s}A^*$. The operator $A_s^*A : X_s \rightarrow X_s$ is self-adjoint. Further, there holds

$$G^s g_n(T^*T) = g_n(A_s^*A)G^s. \quad (4.22)$$

Consequently, the regularized solution (2.1) which is an element of the space X_s can be written in the equivalent form

$$x_n^\delta - x_0 = g_n(A_s^*A)A_s^*(y^\delta - Ax_0).$$

From the valid identity $x_n^\delta - x^\dagger = -r_n(A_s^*A)(x^\dagger - x_0) + g_n(A_s^*A)A_s^*(y^\delta - Ax^\dagger)$ and the identity $g_n(A_s^*A)A_s^* = A_s^*g_n(AA_s^*)$ we obtain

$$\begin{aligned} \|x_n^\delta - x^\dagger\|_s^2 &= \|r_n(A_s^*A)(x^\dagger - x_0)\|_s^2 + \|g_n(A_s^*A)A_s^*(y^\delta - Ax^\dagger)\|_s^2 \\ &\quad - 2(Ag_n(A_s^*A)r_n(A_s^*A)(x^\dagger - x_0), y^\delta - Ax^\dagger). \end{aligned} \quad (4.23)$$

We introduce the abbreviations

$$R_n := g_n(AA_s^*)r_n(AA_s^*) \quad \text{and} \quad y_0^\delta := y^\delta - Ax_0,$$

decompose y_0^δ into the sum $A(x^\dagger - x_0)$ plus $y^\delta - Ax^\dagger$ and obtain the equality

$$\begin{aligned} (R_n y_0^\delta, y_0^\delta) &= (R_n A(x^\dagger - x_0), A(x^\dagger - x_0)) + (R_n (Ax^\dagger - y^\delta), Ax^\dagger - y^\delta) \\ &\quad + 2(R_n A(x^\dagger - x_0), y^\delta - Ax^\dagger). \end{aligned} \quad (4.24)$$

Addition of the equations (4.23) and (4.24) yields

$$\begin{aligned} (R_n y_0^\delta, y_0^\delta) + \|x_n^\delta - x^\dagger\|_s^2 &= \|r_n(A_s^*A)(x^\dagger - x_0)\|_s^2 + \|g_n(A_s^*A)A_s^*(y^\delta - Ax^\dagger)\|_s^2 \\ &\quad + (A_s^*R_n A(x^\dagger - x_0), (x^\dagger - x_0))_s \\ &\quad + (R_n (Ax^\dagger - y^\delta), Ax^\dagger - y^\delta). \end{aligned}$$

We use the valid identities

$$r_n^2(A_s^*A) + A_s^*R_n A = r_n(A_s^*A), \quad Ag_n^2(A_s^*A)A_s^* + R_n = g_n(AA_s^*)$$

and obtain from the above equation

$$\begin{aligned} (R_n y_0^\delta, y_0^\delta) + \|x_n^\delta - x^\dagger\|_s^2 &= (r_n(A_s^*A)(x^\dagger - x_0), x^\dagger - x_0)_s \\ &\quad + (g_n(AA_s^*)(Ax^\dagger - y^\delta), Ax^\dagger - y^\delta). \end{aligned} \quad (4.25)$$

By exploiting properties (i) and (iv) of Proposition 2.1, we obtain

$$\begin{aligned} \text{(a)} \quad \sigma_n^{-1}(g_n(AA_s^*)(Ax^\dagger - y^\delta), Ax^\dagger - y^\delta) &\leq \|Ax^\dagger - y^\delta\|^2, \\ \text{(b)} \quad \sigma_n^{-1}(R_n y_0^\delta, y_0^\delta) &\geq (r_n^2(AA_s^*)y_0^\delta, y_0^\delta) = \|r_n(AA_s^*)y_0^\delta\|^2 = \|Ax_n^\delta - y^\delta\|^2. \end{aligned}$$

We multiply (4.25) by σ_n^{-1} , use the estimates (a) and (b) and obtain

$$\|Ax_n^\delta - y^\delta\|^2 + \sigma_n^{-1}\|x_n^\delta - x^\dagger\|_s^2 \leq \|Ax^\dagger - y^\delta\|^2 + \sigma_n^{-1}(r_n(A_s^*A)(x^\dagger - x_0), x^\dagger - x_0)_s. \quad (4.26)$$

Finally we observe that due to (4.22) we have

$$(r_n(A_s^*A)(x^\dagger - x_0), x^\dagger - x_0)_s = \|r_n^{1/2}(T^*T)G^{-s}(x^\dagger - x_0)\|^2.$$

From this identity and (4.26) we obtain (4.21). \square

From Proposition 4.7 and Proposition 2.8 we obtain the main result of this subsection.

Theorem 4.8. *Let x_n^δ be defined by (2.1) and σ_n be chosen by the discrepancy principle (4.2) with $1 \leq C_1 \leq C_2$, assume the link condition A1(i), the solution smoothness A2 and $0 \leq s \leq p$. If f defined by (2.20) is convex, and $\xi_s(t) := \psi_s^2(t^{1/(2s)})$ is convex where ψ_s is given by (2.12), then*

$$\|x_n^\delta - x^\dagger\| \leq E \left[\psi_p^{-1} \left(\frac{(C_2 + 1)\delta}{mE} \right) \right]^p. \quad (4.27)$$

Proof. For σ_n chosen by the discrepancy principle (4.2) the estimate (4.21) of Proposition 4.7 attains the form

$$C_1^2 \delta^2 + \sigma_n^{-1} \|x_n^\delta - x^\dagger\|_s^2 \leq \delta^2 + \sigma_n^{-1} \|r_n^{1/2}(T^*T)G^{-s}(x^\dagger - x_0)\|^2.$$

Since $C_1 \geq 1$, we have $\|x_n^\delta - x^\dagger\|_s^2 \leq \|r_n^{1/2}(T^*T)G^{-s}(x^\dagger - x_0)\|^2$. We use the representation $G^{-s}(x^\dagger - x_n) = r_n(T^*T)G^{-s}(x^\dagger - x_0)$, see (2.6), use Assumption A2 and obtain

$$\begin{aligned} \|x_n^\delta - x^\dagger\|_s^2 &\leq \|r_n^{1/2}(T^*T)G^{-s}(x^\dagger - x_0)\|^2 \\ &= (G^{p-2s}(x^\dagger - x_n), G^{-p}(x^\dagger - x_0)) \\ &\leq E \|x_n - x^\dagger\|_{2s-p}, \end{aligned} \quad (4.28)$$

where x_n is the regularized solution with exact data. For estimating $\|x_n - x^\dagger\|_{2s-p}$, we use estimate (2.21) of Proposition 2.8 and obtain

$$\|x_n^\delta - x^\dagger\|_s \leq E \left[\psi_p^{-1} \left(\frac{\|\varrho(G)(x_n - x^\dagger)\|}{E} \right) \right]^{p-s}. \quad (4.29)$$

For estimating $\|Ax_n - Ax^\dagger\|$, we use (2.8), the identity $r_n(TT^*)(y^\delta - Ax_0) = y^\delta - Ax_n^\delta$, $r_n(\lambda) \leq 1$ and (4.2) and obtain the estimate

$$\begin{aligned} \|Ax_n - Ax^\dagger\| &= \|r_n(TT^*)(y - Ax_0)\| \\ &\leq \|r_n(TT^*)(y^\delta - Ax_0)\| + \|r_n(TT^*)(y - y^\delta)\| \\ &\leq (C_2 + 1)\delta. \end{aligned}$$

Hence, by using A1(i) we have $\|\varrho(G)(x_n - x^\dagger)\| \leq (C_2 + 1)\delta/m$. Since ψ_p^{-1} is monotone, we obtain from (4.29) the estimate

$$\|x_n^\delta - x^\dagger\|_s \leq E \left[\psi_p^{-1} \left(\frac{(C_2 + 1)\delta}{mE} \right) \right]^{p-s}. \quad (4.30)$$

Next, let us estimate $\|\varrho(G)(x_n^\delta - x^\dagger)\|$. Using Assumption A1(i) and the estimate $\|Ax_n^\delta - Ax^\dagger\| \leq \|Ax_n^\delta - y^\delta\| + \|y - y^\delta\| \leq (C_2 + 1)\delta$ yields

$$\|\varrho(G)(x_n^\delta - x^\dagger)\| \leq (C_2 + 1)\delta/m. \quad (4.31)$$

Now we apply the interpolation estimate (2.19) of Proposition 2.7 and obtain by using (4.30), (4.31) and the abbreviation $\delta_1 := \frac{(C_2+1)\delta}{mE}$ that

$$\|x_n^\delta - x^\dagger\| \leq E [\psi_p^{-1}(\delta_1)]^{p-s} \cdot [\psi_s^{-1}(\delta_1 [\psi_p^{-1}(\delta_1)]^{s-p})]^s. \quad (4.32)$$

From the first equation of (3.24) we have $\psi_s^{-1}(\psi_p(\lambda) \cdot \lambda^{s-p}) = \lambda$. Substituting $\lambda = \psi_p^{-1}(\delta_1)$ yields $\psi_s^{-1}(\delta_1 [\psi_p^{-1}(\delta_1)]^{s-p}) = \psi_p^{-1}(\delta_1)$. From this equation and (4.32) we obtain (4.27). \square

5. Practical implementation

For the practical application of implicit iteration methods in Hilbert scales one has to make different decisions: First, one has to choose the operator B , second, one has to fix the number s in the method (1.2), third, one has to choose the starting value x_0 and to fix the numbers α_k , $k = 1, \dots, n$, and fourth, one has effectively to realize the discrepancy principle (4.1) with a little number n of iteration steps. The choice of B and x_0 should be done in dependence on the expected smoothness of the element $x^\dagger - x_0$ such that Assumption A2 holds true for p sufficiently large, and s should have the magnitude of p . In our further study we concentrate on the choice of the numbers α_k , $k = 1, \dots, n$ for effectively realizing the discrepancy principle (4.1) or (4.2) or (4.3), respectively, with a little number n of iteration steps. In a first proposition we give an upper bound for the regularization parameter of the discrepancy principle in case $n = 1$ which will serve us as starting value for the iteration (1.2). To our best knowledge, so far there have not been upper bounds for the regularization parameter of the discrepancy principle in the literature.

Proposition 5.1. *Let $n = 1$, let x_1^δ the regularized solution (2.1) and let $\alpha_1 = \alpha_D$ be chosen by the discrepancy principle (4.1) with $C \geq 1$. If $\|y^\delta - Ax_0\| > C\delta$, then*

$$\alpha_D < \frac{C\delta \|G^s A^*(y^\delta - Ax_0)\|^2}{(\|y^\delta - Ax_0\| - C\delta) \|y^\delta - Ax_0\|^2}. \quad (5.1)$$

Proof. Let $x_\alpha^\delta = x_0 - (A^*A + \alpha G^{-2s})^{-1} A^*(Ax_0 - y^\delta)$ and $\alpha = \alpha_D$ be the regularization parameter that obeys the discrepancy principle $\|Ax_\alpha^\delta - y^\delta\| = C\delta$. For solving this nonlinear equation, Newton's method applied to the equivalent equation

$$g(r) = \|Ax_{1/r}^\delta - y^\delta\|^{-1} - (C\delta)^{-1} = 0 \quad (5.2)$$

is studied in [16] which results in the iteration

$$r_{k+1} = r_k - \frac{\|Ax_{1/r_k}^\delta - y^\delta\|^{-1} - (C\delta)^{-1}}{r_k^{-3} \left(v_{1/r_k}^\delta, G^{-2s}(x_{1/r_k}^\delta - x_0) \right) \|Ax_{1/r_k}^\delta - y^\delta\|^{-3}} \quad (5.3)$$

where $v_{1/r}^\delta$ is given by $v_{1/r}^\delta = (A^*A + r^{-1}G^{-2s})^{-1} G^{-2s}(x_{1/r}^\delta - x_0)$. From [16, Theorem 3.5] we know that the iteration (5.3) possesses the following properties:

- (i) The sequence (r_k) converges globally and monotonically from the left to r_D for any starting values $0 \leq r_0 < r_D := 1/\alpha_D$.
- (ii) The speed of convergence is locally quadratic.

For $r \rightarrow 0$, the both limit relations

$$\lim_{r \rightarrow 0+0} x_{1/r}^\delta = x_0 \quad \text{and} \quad \lim_{r \rightarrow 0+0} r_k^{-3} (v_{1/r}^\delta, G^{-2s}(x_{1/r}^\delta - x_0)) = \|G^s A^*(y^\delta - Ax_0)\|^2$$

are valid. We execute one iteration step of the iteration (5.3) with starting value $r_0 = 0$ and obtain due to the above limit relations that

$$r_1 = \frac{(\|y^\delta - Ax_0\| - C\delta) \|y^\delta - Ax_0\|^2}{C\delta \|G^s A^*(y^\delta - Ax_0)\|^2}.$$

Due to the above property (i) we have $r_1 < r_D$. Since r and α are related by $\alpha = 1/r$ we obtain (5.1). \square

Based on the Newton iteration (5.3) we propose following strategy for effectively realizing the discrepancy principle (4.3) with a little number n of iteration steps.

Algorithm 1 Global convergent Newton iteration for rule (4.3)

- 1: Start with initial data $y^\delta, A, G, s, \delta, C := 1.1$ and x_0 .
 - 2: **if** $\|Ax_0 - y^\delta\| > C\delta$ **then**
 - 3: Compute α by the right hand side of (5.1) with $C = 1$.
 - 4: Compute $x := x_0 - (A^*A + \alpha G^{-2s})^{-1} A^*(Ax_0 - y^\delta)$ and set $n := 1$.
 - 5: **while** $\|Ax - y^\delta\| > C\delta$ **do**
 - 6: Compute $v := (A^*A + \alpha G^{-2s})^{-1} G^{-2s}(x - x_0)$.
 - 7: Update $r := \frac{1}{\alpha} - \frac{\|Ax - y^\delta\|^{-1} - \delta^{-1}}{\alpha^3 (v, G^{-2s}(x - x_0)) \|Ax - y^\delta\|^{-3}}, n := n + 1,$
 - 8: $x_0 := x, \alpha := 1/r, x := x_0 - (A^*A + \alpha G^{-2s})^{-1} A^*(Ax_0 - y^\delta).$
 - 9: **end while**
 - 10: **end if**
-

For discussing some properties of Algorithm 1, we will work with the notation

$$x_k^\delta(\alpha) := x_{k-1}^\delta - (A^*A + \alpha B^{2s})^{-1} A^*(Ax_{k-1}^\delta - y^\delta), \quad k = 1, 2, \dots,$$

that indicates the dependence of x_k^δ defined by (1.2) on the parameter α . We start by some monotonicity property of the sequence $(\alpha_k)_{k=1}^n$ in the iteration (1.2).

Proposition 5.2. *The regularized solutions $x_k^\delta, k = 1, \dots, n$, obtained by Algorithm 1 have the form (1.2). The related sequence $(\alpha_k)_{k=1}^n$ is strictly monotonically decreasing.*

Proof. In steps 3 and 4 of Algorithm 1, α_1 and $x_1^\delta = x_1^\delta(\alpha_1)$ are computed. Then, the while loop (steps 5 – 9 of Algorithm 1) is executed $n - 1$ times to obtain α_k and $x_k^\delta = x_k^\delta(\alpha_k)$ for $k = 2, \dots, n$. The parameter $\alpha = \alpha_k := 1/r_k$ (see step 7 of Algorithm 1) is obtained by performing one Newton step for solving the nonlinear equation

$$g(r) = \|Ax_{k-1}^\delta(1/r) - y^\delta\|^{-1} - \delta^{-1} = 0$$

with starting value $r_{k-1} = 1/\alpha_{k-1}$. It can be shown (compare [16]) that the function g possesses following properties:

(i) There hold the two limit relations

$$\lim_{r \rightarrow 0+0} g(r) = \|Ax_{k-2}^\delta - y^\delta\|^{-1} - \delta^{-1} < 0 \quad \text{and} \quad \lim_{r \rightarrow \infty} g(r) = +\infty.$$

(ii) The function $g : \mathbb{R}^+ \rightarrow \mathbb{R}$ is monotonically increasing and concave.

From these properties and $g(r_{k-1}) < 0$ we conclude that $r_k > r_{k-1}$. It follows that $\alpha_k < \alpha_{k-1}$ for all $k = 2, \dots, n$. \square

For discussing convergence properties of Algorithm 1 we consider Tikhonov regularization

$$x_1^\delta(\beta_m) := x_0 - (A^*A + \beta_m G^{-2s})^{-1} A^*(Ax_0 - y^\delta) \quad (5.4)$$

and assume

(i) $\beta_k = r_k^{-1}$, $k = 2, \dots, m$, is obtained by the iteration

$$r_k = r_{k-1} - \frac{\|Ax_{1/r_{k-1}}^\delta - y^\delta\|^{-1} - \delta^{-1}}{r_{k-1}^{-3} \left(v_{1/r_{k-1}}^\delta, G^{-2s}(x_{1/r_{k-1}}^\delta - x_0) \right) \|Ax_{1/r_{k-1}}^\delta - y^\delta\|^{-3}}, \quad (5.5)$$

where $v_{1/r}^\delta$ is given by $v_{1/r}^\delta = (A^*A + r^{-1}G^{-2s})^{-1}G^{-2s}(x_{1/r}^\delta - x_0)$ and $x_{1/r}^\delta$ is given by $x_{1/r}^\delta = x_1^\delta(1/r)$,

(ii) r_1 is chosen as $r_1 = \frac{(\|y^\delta - Ax_0\| - \delta)\|y^\delta - Ax_0\|^2}{\delta \|G^s A^*(y^\delta - Ax_0)\|^2}$ and the iteration (5.5)

is stopped with the first integer m for which, with $C := 1.1$,

$$\|Ax_1^\delta(\beta_m) - y^\delta\| \leq C\delta < \|Ax_1^\delta(\beta_k) - y^\delta\|, \quad 0 \leq k < m. \quad (5.6)$$

From [16] we know that the iteration (5.5) converges globally and monotonically from the left to the solution of the equation $g(r) = \|Ax_1^\delta(1/r) - y^\delta\|^{-1} - \delta^{-1} = 0$, and that in the vicinity of the solution we have quadratic speed of convergence. It follows that by the stopping rule (5.6) a finite number m of iteration steps is defined. Our next proposition tells us that Algorithm 1 is not slower than the iteration (5.5) with stopping rule (5.6).

Proposition 5.3. *Let m be the number of iterations of method (5.5) with stopping rule (5.6). Then, $n \leq m$, where n is the number of iterations of Algorithm 1.*

Proof. Assume that α_1 and $x_1^\delta(\alpha_1)$ in steps 3 and 4 of Algorithm 1 are computed, which coincide with β_1 and $x_1^\delta(\beta_1)$ of the iteration (5.5). Then, in the first iteration step of the while-loop (steps 5 – 9 of Algorithm 1) we obtain α_2 and

$$x_2^\delta(\alpha_2) = x_1^\delta(\alpha_1) - (A^*A + \alpha_2 G^{-2s})^{-1} A^*(Ax_1^\delta(\alpha_1) - y^\delta).$$

For $x_2^\delta(\alpha_2)$ computed in this way we have

$$y^\delta - Ax_2^\delta(\alpha_2) = \alpha_2(TT^* + \alpha_2 I)^{-1} \alpha_1(TT^* + \alpha_1 I)^{-1} (y^\delta - Ax_0). \quad (5.7)$$

On the other hand, from the iteration (5.5) we obtain after the first step the regularization parameter $\beta_2 = \alpha_2$ and the regularized solution $x_1^\delta(\beta_2)$ which obeys

$$y^\delta - Ax_1^\delta(\beta_2) = \beta_2(TT^* + \beta_2)^{-1} (y^\delta - Ax_0). \quad (5.8)$$

Comparing both identities (5.7) and (5.8) and observing that $\alpha_2 = \beta_2$ we obtain that $\|y^\delta - Ax_2^\delta(\alpha_2)\| < \|y^\delta - Ax_1^\delta(\beta_2)\|$. In a similar way we obtain that

$$\|y^\delta - Ax_k^\delta(\alpha_k)\| < \|y^\delta - Ax_1^\delta(\beta_k)\|, \quad k = 3, \dots, n,$$

where (α_k) is the sequence generated by Algorithm 1 and (β_k) is the sequence generated by (5.5). From this estimate we obtain that Algorithm 1 terminates not later than the iteration (5.5) with stopping rule (5.6). \square

After termination of Algorithm 1, different cases can appear:

- (1) We have $\delta < \|Ax_n^\delta - y^\delta\| \leq C\delta$ with $C = 1.1$. In this case, all three Theorems 4.5, 4.6 and 4.8 apply.
- (2) We have $\|Ax_n^\delta - y^\delta\| = \delta$. Then, both Theorems 4.6 and 4.8 apply.
- (3) We have $\|Ax_n^\delta - y^\delta\| < \delta$. In this case, Theorem 4.6 applies.

Our next proposition tells us that in all three termination cases (1) – (3), the additional assumption (4.4) of Theorem 4.6 is satisfied with some $c < 1$.

Proposition 5.4. *The regularized solution x_n^δ obtained by Algorithm 1 has the form (2.1) with some sequence $(\alpha_k)_{k=1}^n$ that obeys assumption (4.4) with $c < 1$.*

Proof. Consider the final iteration of the while-loop (steps 5 – 9 of Algorithm 1). Starting from $x = x_{n-1}^\delta(\alpha_{n-1})$ with $\|Ax - y^\delta\| > C\delta$, $\alpha_n := 1/r_n$ is obtained by performing one Newton step for solving the nonlinear equation

$$g(r) = \|Ax_{n-1}^\delta(1/r) - y^\delta\|^{-1} - \delta^{-1} = 0$$

with starting value $r_{n-1} = 1/\alpha_{n-1}$. As a result, we obtain some $\alpha_n < \alpha_{n-1}$, see Proposition 5.2, and the final regularized solution x_n^δ is obtained by

$$x_n^\delta(\alpha_n) = x_{n-1}^\delta(\alpha_{n-1}) - (A^*A + \alpha_n G^{-2s})^{-1} A^* (Ax_{n-1}^\delta(\alpha_{n-1}) - y^\delta).$$

Some formal computations show that $x_n^\delta(\alpha_n)$ can be rewritten as

$$x_n^\delta(\alpha_n) = x_{n-1}^\delta(\alpha_n) - (A^*A + \alpha_{n-1} G^{-2s})^{-1} A^* (Ax_{n-1}^\delta(\alpha_n) - y^\delta).$$

Since the function g is monotonically increasing and concave and since $g(r_{n-1}) < 0$ we conclude that the element $x_{n-1}^\delta(\alpha_n)$ obeys $\|Ax_{n-1}^\delta(\alpha_n) - y^\delta\| > \delta$. It follows that the final two parameters α_{n-1} and α_n in the iteration (1.2) can be interchanged such that we have $\alpha_n > \alpha_{n-1}$. This yields (4.4) with some constant $c < 1$. \square

6. Numerical experiments

In this section we perform numerical experiments for computing regularized solutions by Algorithm 1. We consider Fredholm integral equations

$$[Ax](s) := \int_0^1 K(s, t)x(t) dt = y(s), \quad 0 \leq s \leq 1, \quad A : L^2(0, 1) \rightarrow L^2(0, 1) \quad (6.1)$$

and differential operators $B : D \subset L^2(0, 1) \rightarrow L^2(0, 1)$ of first order defined by

$$Bx = \sum_{k=1}^{\infty} k(x, e_k)e_k \quad \text{with} \quad e_k(t) = \sqrt{2} \sin(k\pi t). \quad (6.2)$$

Example 6.1. Our test example (*deriv2* from [8]) is (6.1) with kernel function

$$K(s, t) = \begin{cases} s(1-t) & \text{for } s \leq t \\ t(1-s) & \text{for } s \geq t. \end{cases}$$

For this kernel function, Assumption A1 is satisfied with $m = M = \pi^{-2}$ and $\varrho(t) = t^2$. We consider three subexamples in which the right hand sides $y(s)$, the corresponding solutions $x^\dagger(t)$ and the maximal smoothness parameters p_0 for which Assumption A2 with $x_0 = 0$ holds true for all $p \in (0, p_0)$, are given by

$$\begin{aligned} \text{(i)} \quad & y(s) = -\frac{1}{4\pi^2} \sin 2\pi s, \quad x^\dagger(t) = \sin 2\pi t, \quad p_0 = \infty, \\ \text{(ii)} \quad & y(s) = \frac{s}{3} (1 - 2s^2 + s^3), \quad x^\dagger(t) = 4t(1-t), \quad p_0 = \frac{5}{2}, \\ \text{(iii)} \quad & y(s) = \frac{s}{6} (1 - s^2), \quad x^\dagger(t) = t, \quad p_0 = \frac{1}{2}. \end{aligned}$$

The discretization of (6.1) has been done by Galerkin approximation as outlined, e. g., in [8, 16], guaranteeing that $\|x^\dagger\|_2 \approx \|x^\dagger(t)\|_{L^2(0,1)}$ and $\|y\|_2 \approx \|y(s)\|_{L^2(0,1)}$ holds. As a discrete approximation of the first order differential operator (6.2) we use the (m, m) – matrix

$$B := B_2^{1/2} \quad \text{with} \quad B_2 = \frac{(m+1)^2}{\pi^2} \begin{pmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix}, \quad (6.3)$$

compare [16]. For modeling noise in the discretized right hand side $y \in \mathbb{R}^m$, for given nonnegative σ we compute

$$y^\delta = y + \sigma \frac{\|y\|_2}{\|e\|_2} e,$$

where $e = (e_i)$ is a random vector with $e_i \sim \mathcal{N}(0, 1)$. In this way of modeling noise we guarantee that for the relative error we have $\|y - y^\delta\|_2 / \|y\|_2 = \sigma$. The noise level δ is then given by $\delta = \sigma \|y\|_2$. Tables 1 – 3 show our numerical results with $x_0 = 0$, where the letter codes in the leftmost column refer to following three iteration methods:

- **(TI/DP):** This is the method of Tikhonov regularization (5.4) with $x_0 = 0$ and $s = 1$, where the regularization parameter obeys (5.6) and is obtained by the iteration (5.5) which converges globally and locally quadratically.
- **(IIM/A1):** This is the implicit iteration method (1.2) with $x_0 = 0$ and $s = 1$, where the sequence $(\alpha_k)_{k=1}^n$ is obtained by Algorithm 1.
- **(IIM/GS):** This is the implicit iteration method (1.2) with $x_0 = 0$ and $s = 1$, where the sequence $(\alpha_k)_{k=1}^n$ is the geometric sequence $(q^{k-1}\alpha_1)_{k=1}^n$ with $q = \frac{1}{2}$ as proposed in [7] and stopping rule (4.3) with $C = 1.1$.

For all three iteration methods our tables contain

- (i) the number n of required iterations,
- (ii) the final regularization parameter α_n ,

- (iii) the discrepancy $d_n := \|Ax_1^\delta(\alpha_n) - y^\delta\|_2$ of the final approximation for the iteration method (TI/DP), and the discrepancy $d_n := \|Ax_n^\delta - y^\delta\|_2$ of the final approximation for the iteration methods (IIM/A1) and (IIM/GS),
- (iv) the error $e_n := \|x_1^\delta(\alpha_n) - x^\dagger\|_2$ of the final approximation for the iteration method (TI/DP), and the error $e_n := \|x_n^\delta - x^\dagger\|_2$ of the final approximation for the iteration methods (IIM/A1) and (IIM/GS).

In our experiments, all three iteration methods have been started first with

$$\alpha_1 = \delta \frac{(B^{-2}A^*y^\delta, A^*y^\delta)}{(\|y^\delta\|_2 - \delta) \|y^\delta\|_2^2}, \quad (6.4)$$

compare (5.1), and second with $\alpha_1 = 1$ as done in [7]. In order to keep the discretization error small, we have used the dimension number $m = 400$ in all computations.

Method	n	α_n	d_n	e_n
(TI/DP)	3	5.54 E-7	1.88 E-4	3.79 E-3
(IIM/A1)	2	8.85 E-7	1.78 E-4	2.94 E-3
(IIM/GS)	2	8.10 E-7	1.78 E-4	2.95 E-3
(TI/DP)	4	5.54 E-7	1.88 E-4	3.79 E-3
(IIM/A1)	3	8.85 E-7	1.78 E-4	2.94 E-3
(IIM/GS)	17	1.52 E-5	1.78 E-4	2.83 E-3

TABLE 1. Example 6.1 (i) with $\sigma = 0.01$ ($\delta = \sigma\|y\|_2 \approx 1.79$ E-4). *Top:* α_1 from (6.4), *Down:* $\alpha_1 = 1$.

Method	n	α_n	d_n	e_n
(TI/DP)	3	2.75 E-5	7.97 E-4	1.62 E-2
(IIM/A1)	2	5.14 E-5	7.75 E-4	1.70 E-2
(IIM/GS)	2	5.18 E-5	7.75 E-4	1.70 E-2
(TI/DP)	4	2.75 E-5	7.97 E-4	1.62 E-2
(IIM/A1)	3	5.15 E-5	7.75 E-4	1.70 E-2
(IIM/GS)	12	4.88 E-4	8.08 E-4	2.46 E-2

TABLE 2. Example 6.1 (ii) with $\sigma = 0.01$ ($\delta = \sigma\|y\|_2 \approx 7.39$ E-4). *Top:* α_1 from (6.4), *Down:* $\alpha_1 = 1$.

In our numerical experiments we observed that the accuracy of each individual regularization method in the three test cases of Examples 6.1 (i) – (iii) is as predicted by the theory. In Tables 1 – 3 we mainly concentrate on the performance of the three methods and observe following:

- (1) As far as computational expenses are concerned, the iteration method (IIM/A1) performs best. In fact, this method requires the smallest number of iterations compared with the other two methods.

Method	n	α_n	d_n	e_n
(TI/DP)	6	7.62 E−8	4.83 E−4	1.51 E−1
(IIM/A1)	5	1.24 E−7	4.80 E−4	1.51 E−1
(IIM/GS)	9	3.98 E−7	4.95 E−4	1.62 E−1
(TI/DP)	7	7.62 E−8	4.83 E−4	1.51 E−1
(IIM/A1)	6	1.24 E−7	4.80 E−4	1.51 E−1
(IIM/GS)	22	4.76 E−7	5.00 E−4	1.65 E−1

TABLE 3. Example 6.1 (iii) with $\sigma = 0.01$ ($\delta = \sigma\|y\|_2 \approx 4.60$ E−4). *Top:* α_1 from (6.4), *Down:* $\alpha_1 = 1$.

- (2) For the method (IIM/GS), the number of iterations can considerably be reduced by starting with α_1 from (6.4) instead of starting with $\alpha_1 = 1$. For the other two methods (TI/DP) and (IIM/A1), the number of iterations differs only by 1 for the two starting values (6.4) and $\alpha_1 = 1$, respectively.
- (3) In all three iteration methods, the α -sequence $(\alpha_k)_1^n$ is decreasing. However, the final regularization parameter α_n is smallest for method (TI/DP). Comparing the discrepancies d_k for the individual iterations $k = 1, 2, \dots$ (which are not contained in the tables) we observed that, for $k \geq 2$, d_k in method (IIM/A1) is always smaller than d_k in method (TI/DP).

References

- [1] P. Bégout and F. Soria, *A generalized interpolation inequality and its application to the stabilization of damped equations*, J. Diff. Equ. 240 (2007), 324–356.
- [2] R. Bhatia, *Matrix Analysis*, Springer, New York, 1997.
- [3] A. Böttcher, B. Hofmann, U. Tautenhahn, and M. Yamamoto, *Convergence rates for Tikhonov regularization from different kinds of smoothness conditions*, Appl. Anal. 85 (2006), 555–578.
- [4] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
- [5] U. Hämarik and U. Tautenhahn, *On the monotone error rule for parameter choice in iterative and continuous regularization methods*, BIT 41 (2001).
- [6] M. Hanke, *The regularizing Levenberg–Marquardt scheme is of optimal order*, J. Integral Equations Appl. 22 (2010), 259–283.
- [7] M. Hanke and C. W. Groetsch, *Nonstationary iterated Tikhonov regularization*, J. Optim. Theory Appl. 97 (1998), 37–53.
- [8] P. C. Hansen, *Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems*, Numerical Algorithms 6 (1994), 1–35.
- [9] M. Hegland, *An optimal order regularization method which does not use additional smoothness assumptions*, SIAM J. Numer. Anal. 5 (1992), 1446–1461.
- [10] M. Hegland, *Variable Hilbert scales and their interpolation inequalities with application to Tikhonov regularization*, Appl. Anal. 59 (1995), 207–223.
- [11] M. Hochbruck, M. Hönl, and A. Ostermann, *A convergence analysis of the exponential Euler iteration for nonlinear ill-posed problems*, Inverse Problems 25 (2009), 075009 (18pp).
- [12] T. Hohage and M. Pricop, *Nonlinear Tikhonov regularization in Hilbert scales for inverse boundary value problems with random noise*, Inverse Problems and Imaging 2 (2008), 271–290.

- [13] Q. Jin, *On a regularized Levenberg-Marquardt method for solving nonlinear ill-posed problems*, Numer. Math. 115 (2010), 229–259.
- [14] A. Lechleiter and A. Rieder, *Towards a general convergence theory for inexact Newton regularizations*, Numer. Math. 114 (2009), 521–541.
- [15] F. Liu and M. Z. Nashed, *Tikhonov regularization of nonlinear ill-posed problems with closed operators in Hilbert scales*, J. Inv. Ill-Posed Problems 5 (1997), 363–376.
- [16] S. Lu, S. V. Pereverzev, S. Shao, and U. Tautenhahn, *On the generalized discrepancy principle for Tikhonov regularization in Hilbert scales*, J. Integral Equations Appl. (accepted) (2010).
- [17] B. A. Mair, *Tikhonov regularization for finitely and infinitely smoothing operators*, SIAM J. Math. Anal. 25 (1994), 135–147.
- [18] P. Mathé and B. Hofmann, *Direct and inverse results in variable Hilbert scales*, J. Approx. Theory 154 (2008), 77–89.
- [19] P. Mathé and S. V. Pereverzev, *Geometry of linear ill-posed problems in variable Hilbert scales*, Inverse Problems 19 (2003), 789–803.
- [20] P. Mathé and S. V. Pereverzev, *Moduli of continuity for operator valued functions*, Num. Funct. Anal. and Optimiz. 23 (2003), 623–631.
- [21] P. Mathé and U. Tautenhahn, *Interpolation in variable Hilbert scales with application to inverse problems*, Inverse Problems 22 (2006), 2271–2297.
- [22] P. Mathé and U. Tautenhahn, *Error bounds for regularization methods in Hilbert scales by using operator monotonicity*, Far East J. Math. Sci. 24 (2007), 1–21.
- [23] V. A. Mikhailets and A. A. Murach, *Interpolation with a function parameter and refined scale of spaces*, Methods Funct. Anal. Topology 14 (2008), 81–100.
- [24] V. A. Morozov, *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl. 7 (1966), 414–417.
- [25] M. T. Nair, S. V. Pereverzev, and U. Tautenhahn, *Regularization in Hilbert scales under general smoothing conditions*, Inverse Problems 21 (2005), 1851–1869.
- [26] M. T. Nair, E. Schock, and U. Tautenhahn, *Morozov’s discrepancy principle under general source conditions*, Z. Anal. Anw. 22 (2003), 199–214.
- [27] M. T. Nair and U. Tautenhahn, *Convergence rates for Lavrentiev-type regularization in Hilbert scales*, Comput. Meth. Appl. Math. 8 (2008), 279–293.
- [28] F. Natterer, *Error bounds for Tikhonov regularization in Hilbert scales*, Appl. Anal. 18 (1984), 29–37.
- [29] A. Neubauer, *An a-posteriori parameter choice for Tikhonov-regularization in Hilbert scales leading to optimal convergence rates*, SIAM J. Numer. Anal. 25 (1988), 1313–1326.
- [30] A. Neubauer, *Tikhonov regularization of nonlinear ill-posed problems in Hilbert scales*, Appl. Anal. 46 (1992), 59–72.
- [31] P. Pornsawad and C. Böckmann, *Convergence rate analysis of the first-stage Runge-Kutta-type regularizations*, Inverse Problems 26 (2010), 035005 (12pp).
- [32] T. Reginska and U. Tautenhahn, *Conditional stability estimates and regularization with applications to Cauchy problems for the Helmholtz equation*, Num. Funct. Anal. and Optimiz. 30 (2009), 1065–1097.
- [33] T. Schröter and U. Tautenhahn, *Error estimates for Tikhonov regularization in Hilbert scales*, Num. Funct. Anal. and Optimiz. 15 (1994), 155–168.
- [34] U. Tautenhahn, *Optimal parameter choice for Tikhonov regularization in Hilbert scales*, Inverse Problems in Mathematical Physics, Lecture Notes in Phys. 422 (Berlin) (L. Päiväranta and E. Somersalo, eds.), Springer, 1993, pp. 242–250.
- [35] U. Tautenhahn, *Error estimates for regularization methods in Hilbert scales*, SIAM J. Numer. Anal. 33 (1996), 2120–2130.
- [36] U. Tautenhahn, *On a general regularization scheme for nonlinear ill-posed problems: II. Regularization in Hilbert scales*, Inverse Problems 14 (1998), 1607–1616.
- [37] U. Tautenhahn, *Optimality for linear ill-posed problems under general source conditions*, Num. Funct. Anal. and Optimiz. 19 (1998), 377–398.

- [38] M. Uchiyama, *Majorization and some operator monotone functions*, Linear Algebra and its Applications 432 (2010), 1867–1872.
- [39] G. M. Vainikko and A. Y. Veretennikov, *Iteration Procedures in Ill-Posed Problems*, Nauka, Moscow, 1986, In Russian.

Q. JIN, DEPARTMENT OF MATHEMATICS, VIRGINIA TECH BLACKSBURG, 522 MCBRYDE HALL, VA 24061-0123, U.S.A.

E-mail address: `qnjin@math.vt.edu`

U. TAUTENHAHN, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF APPLIED SCIENCES ZITTAU/GÖRLITZ, P.O.BOX 1455, 02755 ZITTAU, GERMANY

E-mail address: `u.tautenhahn@hs-zigr.de`